Lectures Notes on measure theory and integration theory as well as differentiation theory (courses TMV100 / MMA110) by Jeffrey Steif (COMMENTS WELCOMED)

September 30, 2021

Contents

Part I

1 Background

While of course one needs to have various prerequisites for this course, I only mention here one thing which you might not have seen. That is the notion of a lim sup and lim inf of a sequence which does not necessarily converge. This is described in Chapter 0 of Folland, which one should look through.

2 Motivation

It turns out that Riemann integration is insufficient for many purposes; it does not behave well with respect to limiting operations. Lebesgue invented in 1902 what would be called Lebesgue measure which will play a central role here.
Goal: We want to assign a "length" or a "size" ℓ to *all* subsets of R satisfying the following three reasonable properties.
1. If $A = [a, b]$, then $\ell(A) = b - a$. (ℓ should correspond to length for intervals.)
2. If $A_1, A_2, \ldots$ are disjoint sets, then

$$\ell(\bigcup_i A_i) = \sum_i \ell(A_i).$$

(The size of a disjoint union of sets is the sum of the sizes of the pieces.)
3. For all sets $A \subseteq R$ and $x \in R$,

$$\ell(A + x) = \ell(A).$$

This is called translation invariance and just means if you shift a set to the right or left, its size should not change. $A+x$ is "A translated by x"; formally $A + x = \{a + x : a \in A\}$.

This is a natural thing to try and hope for. Unfortunately, we will see that such a length notion does not exist, **assuming the axiom of choice**. We will need to drop some assumption. What we will do is to drop the assumption that ℓ is defined for ALL subsets. However, it will be defined for all subsets you could ever imagine and it will satisfy 1-3 for them. This will then be called *Lebesgue measure*.

The following question is some motivation for the theory of Lebesgue integration and demonstrates the subtleties that arise with elementary limiting operations.

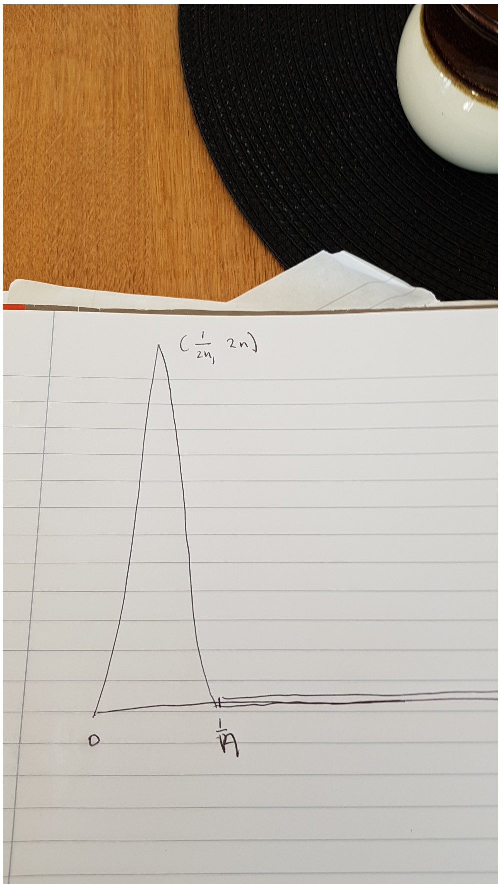
Question: If f_n is a nonnegative continuous function on $[0, 1]$ for each n and if

$$\lim_{n\to\infty} f_n(x) = 0$$

for all $x \in [0, 1]$ (we say f_n goes to 0 pointwise in this case), does it follow that

$$\lim_{n\to\infty} \int_0^1 f_n(x)dx = 0?$$

It seems natural that the answer is yes, but in fact it is false. Let f_n be a "sharply pointed tent" function which starts at 0 at 0, increases linearly at rate $4n^2$ until $1/(2n)$ (at which point the function has value $2n$) and then decreases linearly at rate $4n^2$ until $1/n$ where it hits 0 and then stays at 0 on $[1/n, 1]$. (See picture.)



Exercise: Verify that this sequence works; i.e., f_n goes to 0 pointwise but each of the integrals is 1.

Question: When *will* we be able to conclude from the fact that f_n goes to 0 pointwise that the integrals converge to 0?

We will see an answer to this question later on.

Another motivation for Lebesgue integration, which we will not go into detail in these notes, but will just mention here is the following.

First, let us recall how the real numbers arise. The integers are a natural construction and from these it is natural that the rationals arise simply because of division. Now it is difficult to do "analysis" with the rationals since it has "lots of holes". More rigorously, it is not a complete metric space and we obtain the real numbers by "filling in these holes" or more rigorously, by taking the completion of the metric space of rationals. The completeness of the real numbers is a crucial property which allows us to do "analysis". For example, we would often not have solutions to $x^2 = a$ for $a > 0$ if we did not do this completion.

Now consider the space of continuous functions on $[0, 1]$ together with the metric

$$d(f, g) := \int_0^1 |f(x) - g(x)| \, dx.$$

It it easy to find Cauchy sequences in this metric space which don't converge. Therefore, just as we completed the rationals to obtain the real numbers, it is natural to take this metric space and complete it (as a metric space). When one does this, one obtains essentially what will be called the space of "measurable functions" and the Lebesgue integral which we will get to later on.

3 Measure Theory

3.1 Algebras and σ-algebras

Question: What properties should we expect for the collection of subsets of R or of $[0,1]$ to which we will assign a "length"?

Definition 3.1. *Let X be a nonempty set. An* **algebra** *or* **field** *of subsets of X is a collection $\mathcal{A}$ of subsets of X which is "closed under finite set theoretic operations"; i.e.*
(1). $X \in \mathcal{A}$, $\emptyset \in \mathcal{A}$
(2). $A_1, A_2, \ldots, A_n$ each in $\mathcal{A}$ implies that $\bigcup_{i=1}^{n} A_i \in \mathcal{A}$ ($\mathcal{A}$ is closed under **finite** *unions)*
(3). $A \in \mathcal{A}$ implies that $A^c \in \mathcal{A}$ ($\mathcal{A}$ is closed under complementation)

Exercise: Show (a) that it is enough to assume that the property in (2) holds for $n = 2$ to conclude it for all n and (b) that (2) and (3) imply that $\mathcal{A}$ is closed under finite intersections.

Definition 3.2. *Let X be a nonempty set. A* **σ-algebra** *or* **σ-field** *of subsets of X is a collection $\mathcal{M}$ of subsets of X which is an algebra and in addition, (2) above is replaced by the stronger*
(2'). $A_1, A_2, \ldots$ each in $\mathcal{M}$ implies that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{M}$ ($\mathcal{M}$ is closed under countable *unions)*

So, in words, a σ-algebra is closed under countable set theoretic operations.

Exercise: Show that a σ-algebra is also closed under finite unions.
Exercise: Show that a σ-algebra is also closed under countable intersections.
Exercise: Find an algebra which is not a σ-algebra.
Exercise: Find all algebras on X where $X = \{1, 2, 3\}$.

Remark: The number of σ-algebras on a set of size n are called the Bell numbers and are much studied in combinatorics (motivated by other things than σ-algebras). These numbers are known to grow very rapidly in n, faster than any exponential function.

Here are three examples of σ-algebras on an arbitrary set X.

Example 3.3.

1. $\mathcal{F}_1 = \{X, \emptyset\}$
2. $\mathcal{F}_2 = \{A : A$ *is countable or* A^c *is countable*$\}$ *(finite sets are considered countable)*
3. $\mathcal{F}_3$ *consists of all subsets of* X *(called the powerset of* X*, denoted* $\mathcal{P}(X)$*).*

Remark: If A^c is countable, we say that A is *cocountable.*

Exercises.
1. Verify that each of these examples are a σ-algebra.
2. For which X does one have $\mathcal{F}_1 = \mathcal{F}_2$?
3. For which X does one have $\mathcal{F}_2 = \mathcal{F}_3$? (This might be harder than it sounds. Think about it at your own risk.)

Proposition 3.4. *Given a collection* $\mathcal{E}$ *of subsets of* X *(i.e., a subset of* $\mathcal{P}(X)$*), there is a smallest* σ*-algebra containing* $\mathcal{E}$*, denoted by* $\sigma(\mathcal{E})$*, called the* σ*-algebra generated by* $\mathcal{E}$*.*

Proof:
Consider

$$\sigma(\mathcal{E}) := \bigcap_{\mathcal{F}\supseteq\mathcal{E}:\mathcal{F} \text{ is a } \sigma\text{-algebra}} \mathcal{F}.$$

This is the same as

$$\{A : A \text{ is an element of every } \sigma\text{-algebra which contains } \mathcal{E}\}.$$

1. This is a nonempty intersection since $\mathbb{P}(X) \supseteq \mathcal{E}$
2. $\sigma(\mathcal{E})$ contains $\mathcal{E}$ by construction.
3. $\sigma(\mathcal{E})$ is a σ-algebra (Check this. It is easier than it might look; it is just very elementary set theory).

This is clearly the smallest σ-algebra containing $\mathcal{E}$ since it is, by construction, contained inside of every σ-algebra which contains $\mathcal{E}$.
QED

Remarks: The construction above is identical to similar constructions one makes in linear algebra or group theory. Eg., the subgroup generated by a bunch of elements in a group is the intersection of all subgroups containing those elements. Similar for vector spaces.

(THIS IS A LONG **OPTIONAL** ASIDE.)
Remark: We have built the σ-algebra $\sigma(\mathcal{E})$ "from the outside". One may hope that you can build $\sigma(\mathcal{E})$ "from the inside". This is possible but much harder.

You might try to do this as follows. You start with $\mathcal{E}$ and first add all complements of the sets in $\mathcal{E}$ since they would have to be in $\sigma(\mathcal{E})$. (One says "we close under complementation".) Then, starting from there, you take all possible countable unions since they must belong to $\sigma(\mathcal{E})$. (One says "we close under countable unions".) But now you are not closed under complementation and so you have to close again under complements. Now you are not closed under countable unions and so you have to close again under countable unions, etc., etc and you have to "keep going"....forever. And then start again! And it is technical. But it can be done. But we don't here.

It is interesting to compare this with building a topology from a collection of sets which is surprisingly much easier. If I am given a collection $\mathcal{E}$ of subsets of a set X, the topology generated by $\mathcal{E}$ is the collection of all arbitrary unions of finite intersections of sets in $\mathcal{E}$. I.e., you first take all finite intersections of sets in $\mathcal{E}$ since these must be there and then you take arbitrary unions of this latter collection since all these must also be in the generated topology. But now you verify that what you now have **is** a topology and so you are done. (END OF LONG ASIDE.)

Recall the definition of an open set in R: O is open if for all $x \in O$, there exists $\epsilon > 0$ so that $(x - \epsilon, x + \epsilon) \subseteq O$.

Definition 3.5. *The σ-algebra generated by the open subsets of R is called the* Borel σ-algebra of R *and is denoted by $\mathcal{B}$. The sets in $\mathcal{B}$ are called Borel sets.*

Exercise: Show that $\mathcal{B}$ is also (1) the σ-algebra generated by the open intervals and (2) the σ-algebra generated by the closed intervals.

Most sets (and very likely all sets) that you have seen are Borel sets.

There are other notions of sets "being closed under certain operations" which turn out to be useful. Here are two such.

Definition 3.6. *Let X be a nonempty set.*
A nonempty collection $\mathcal{I}$ of subsets of X is called a π-**system** *if it is closed under finite intersections; i.e., $A, B \in \mathcal{I}$ implies $A \cap B \in \mathcal{I}$.*

Definition 3.7. *Let X be a nonempty set.*
A nonempty collection $\mathcal{D}$ of subsets of X is called a $\mathcal{D}$-**system** *if*
a. $X \in \mathcal{D}$
b. $E, F \in \mathcal{D}$ and $E \subseteq F$ imply $F \setminus E (= F \cap E^c) \in \mathcal{D}$
and
c. $E_1 \subseteq E_2 \subseteq E_3, \ldots$ and $E_i \in \mathcal{D}$ for all i imply $\bigcup_i E_i \in \mathcal{D}$.

(Note that a $\mathcal{D}$-system is closed under complementation. Why?)

It is natural to ask why in the world we would introduce such crazy classes of sets. We will see later that they will be very useful.

Theorem 3.8. *(Theorem 3.8 in JJ). If $\mathcal{M}$ is a collection of subsets of a set X, then $\mathcal{M}$ is a σ-algebra if and only if $\mathcal{M}$ is a π-system and a $\mathcal{D}$-system.*

Proof:
The right (only if) implication is immediate which you should check and the left (if) implication is left as a good exercise. (The proof is in JJ).
QED

Given a collection of $\mathcal{E}$ of subsets of X, we have previous defined $\sigma(\mathcal{E})$ as the smallest σ-algebra containing. We do something similar here.

Definition 3.9. *We let $\pi(\mathcal{E})$ ($\mathcal{D}(\mathcal{E})$) be the smallest π-system ($\mathcal{D}$-system) containing $\mathcal{E}$.*

Exercise: Show that these exist and how to construct them. (Hint: The proof is exactly the same as is done for $\sigma(\mathcal{E})$.)

The next theorem is a very important tool. It is called Dynkin's $\pi - \lambda$ Theorem (A $\mathcal{D}$-system is sometimes called a λ-system.) The proof is sort of messy set theory and is not so intuitive. We probably will not do the proof but we certainly will apply it on two occasions. We will see later on why it is useful.

Theorem 3.10. *(Theorem 3.9 in JJ). If $\mathcal{I}$ is a π-system, then*

$$\mathcal{D}(\mathcal{I}) = \sigma(\mathcal{I}) .$$

Note that $\subseteq$ is trivial.

3.2 Some General Measure Theory

Definition 3.11. *If $\mathcal{M}$ is a σ-algebra of subsets of X, then $(X, \mathcal{M})$ is called a* measurable space. *(Not a measure space since there is no measure yet!)*

We will want to define a "size" to each subset which belongs to $\mathcal{M}$ but not to all subsets of X. The following is the crucial definition.

Definition 3.12. *If $(X, \mathcal{M})$ is a measurable space, a* **measure** *m on $(X, \mathcal{M})$ is a mapping from $\mathcal{M}$ to $[0, \infty]$ satisfying the following.*
1. $m(\emptyset) = 0$
2. If $A_1, A_2, \ldots,$ are (pairwise) disjoint elements of $\mathcal{M}$, then

$$m(\bigcup_i A_i) = \sum_i m(A_i).$$

Remarks: 1. The crucial second property is called **countable additivity** and implies **finite additivity** using 1. (Check this.)
2. For $A \in \mathcal{M}$, $m(A)$ is thought of as either the (a) measure, (b) size, (c) length or (d) probability of A.

Definition 3.13. *A measure space $(X, \mathcal{M}, m)$ is a measurable space $(X, \mathcal{M})$ together with a measure m on it.*

Example. Let $X = \{1, 2, 3, \ldots\}$ and consider a vector $p_1, p_2, \ldots$ of nonnegative numbers with $\sum_{i=1}^{\infty} p_i = 1$. Then let $\mathcal{M}$ be all subsets of X and for $S \subseteq X$, let

$$m(S) := \sum_{i \in S} p_i.$$

Exercise. Verify that $(X, \mathcal{M}, m)$ is a measure space. Does it have any probabilistic interpretation?

Theorem 1.8 in Folland gives a number of relatively easy properties of measures.

Theorem 3.14. *Let $(X, \mathcal{M}, m)$ be a measure space.*
a. (Monotonicity) $E, F \in \mathcal{M}$, $E \subseteq F$ implies $m(E) \leq m(F)$.
b. (Continuity from below) $E_1 \subseteq E_2 \subseteq E_3, \ldots$ with each $E_i \in \mathcal{M}$ implies that

$$m(\bigcup_i^{\infty} E_i) = \lim_{n \to \infty} m(E_n).$$

c. (Subadditivity) $E_1, E_2, \ldots \in \mathcal{M}$, then

$$m(\bigcup_i^{\infty} E_i) \leq \sum_i^{\infty} m(E_i).$$

d. (Continuity from above) $E_1 \supseteq E_2 \supseteq E_3, \ldots$ with each $E_i \in \mathcal{M}$ implies that

$$m(\bigcap_i^{\infty} E_i) = \lim_{n \to \infty} m(E_n)$$

provided $m(E_1) < \infty$.

Proof:
a. Using finite additivity in the first step and $m \geq 0$ in second step gives

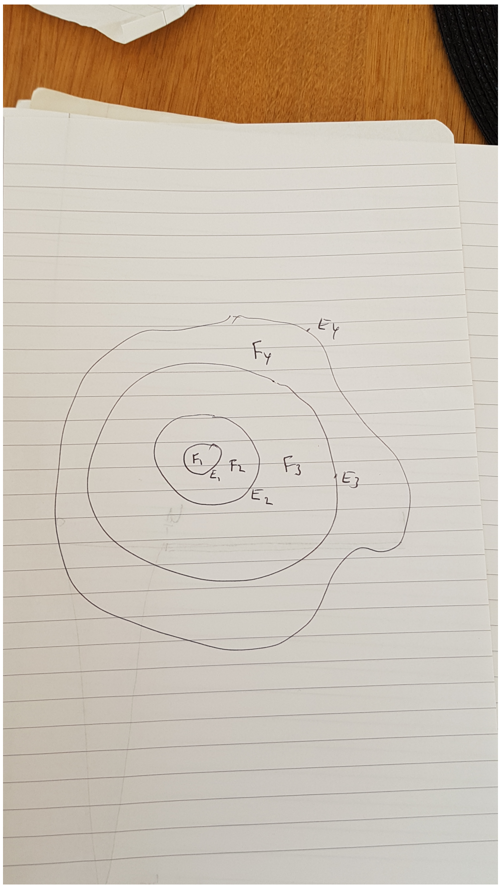
$$m(F) = m(E) + m(F \backslash E) \geq m(E).$$

b. By (a), $m(E_i)$ is a (weakly) increasing sequence and hence the limit exists (possibly ∞ which is fine). Let (see picture)

$$F_1 := E_1, F_2 := E_2 \backslash E_1, F_n := E_n \backslash E_{n-1}, \ldots$$

and observe that (1) the F_i's are disjoint, (2) $E_n = \bigcup_{i=1}^n F_i$ and (3) $\bigcup_{i=1}^\infty E_i = \bigcup_{i=1}^\infty F_i$.
We then have, using countable and finite additivity

$$m(\bigcup_i^\infty E_i) = m(\bigcup_i^\infty F_i) = \sum_i^\infty m(F_i) = \lim_{n\to\infty} \sum_i^n m(F_i) = \lim_{n\to\infty} m(E_n).$$



c. Let

$$F_1 := E_1, F_2 := E_2 \backslash E_1, F_3 := E_3 \backslash (E_1 \cup E_2), \ldots F_n := E_n \backslash (E_1 \cup \ldots \cup E_{n-1}), \ldots$$

and observe that (1) the F_i's are disjoint, (2) $\bigcup_{i=1}^n E_i = \bigcup_{i=1}^n F_i$ and (3) $\bigcup_{i=1}^\infty E_i = \bigcup_{i=1}^\infty F_i$.

We then have

$$m(\bigcup_{i=1}^{\infty} E_i) = m(\bigcup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} m(F_i) \leq \sum_{i=1}^{\infty} m(E_i).$$

d. Exercise.
QED

Exercise: (a). Prove part (d) above. (Hint: Apply the increasing sequence result to the complements and subtract. How are you using the assumption that $m(E_1) < \infty$?)
b. Check (d) is false without the extra assumption. (Hint: Consider Lebesgue measure (length) on R (which we have NOT proved exists yet) and take $E_n = [n, \infty)$.)

Remark: If you are doing probability theory, then this extra assumption will always hold.

Where are we now? We have defined a measure space, proved one theorem about them and constructed one where X is countable. **But, we haven't yet given a single example of an interesting measure space.** Before remedying this situation, we need to give some more definitions.

Definition 3.15. *A measure space* $(X, \mathcal{M}, m)$ *is* **complete** *if (i)* $B \in \mathcal{M}$*, (ii)* $m(B) = 0$ *and (iii)* $A \subseteq B$ *imply that* $A \in \mathcal{M}$ *(which then of course implies that* $m(A) = 0$*).*

Definition 3.16. *Given a measure space* $(X, \mathcal{M}, m)$*, a property (formally a subset of* X*) is said to occur* **almost everywhere** *abbreviated a.e. (***almost surely** *abbreviated a.s. if one is doing probability theory) if the set of* x*'s where the property fails is contained inside of a set of measure 0.*

Remark: If $(X, \mathcal{M}, m)$ were complete, we could have just written that "the set of x's where the property fails has measure 0" but if the space is not complete, "the set of x's where the property fails" might not belong to $\mathcal{M}$ and that is why we needed the definition above.

This issue that a measure space might not be complete will not arise many times at all. It also turns out that you can always "complete a measure space", "adding some sets and extending the measure" and making it complete. We will not discuss this but mention that $(R, \mathcal{B}, m)$, where m is Lebesgue measure, is not complete and when you complete it, you get $(R, \mathcal{M}, m)$ where $\mathcal{M}$ is the collection of "Lebesgue measurable sets" which we have not yet defined.

We end with a few more standard definitions associated to measures.

Definition 3.17. *A measure space* $(X, \mathcal{M}, \mu)$ *is called* **finite** *if* $\mu(X) < \infty$. *(If* $\mu(X) = 1$*, it is called a probability space.)*

Definition 3.18. *A measure space* $(X, \mathcal{M}, \mu)$ *is called* σ**-finite** *if there exist subsets* $A_1, A_2, \ldots$ *so that* $X = \bigcup_i A_i$ *and* $\mu(A_i) < \infty$ *for all* i*.*

Exercises:
1. Show Lebesgue measure on R is σ-finite.
2. Can one take the A_i's in the definition of σ-finiteness to be disjoint?
3. Construct an example of a measure space which is not σ-finite.

Remark: We will see that a number of theorems in this course require the measure space to be σ-finite (which of course includes the finite case).

Definition 3.19. *Assume* $(X, \mathcal{M}, \mu)$ *is a measure space with all single points being measurable. An* **atom** *is a point* x *with* $\mu(\{x\}) > 0$*. Letting* $\mathcal{A}$ *be the set of atoms,* $(X, \mathcal{M}, \mu)$ *is called* **atomic** *if* $\mathcal{A} \in \mathcal{M}$ *and* $\mu(\mathcal{A}^c) = 0$*.* $(X, \mathcal{M}, \mu)$ *is called* **continuous** *if there is no atom.*

Remarks: The definition of an atomic space is sometimes more general than given here but that won't concern us.

Exercise:
If an atomic measure space is σ-finite, then $\mathcal{A}$ must be countable and hence automatically measurable.

3.3 The construction of Lebesgue Measure

We now begin the journey to prove the existence of so-called *Lebesgue measure.*

Theorem 3.20. *There exists a translation invariant measure* m *on* $(R, \mathcal{B})$ *such that* $m([a, b]) = b - a$ *for all* $a < b$*. (*m *will then be Lebesgue measure restricted to* $\mathcal{B}$*.)*

This says we can *almost* do what we originally asked; the difference is that our "length" is not defined for all subsets but all reasonable ones.

People have different approaches to measure theory; does one develop things explictly for Lebesgue measure or more generally? Here we will follow a hybrid which I think works well.

It is important to have the **general plan** or **big picture** on how we construct Lebesgue measure before diving into the details. There will be 5 steps.

STEP 1: Define the general concept of **outer measure**.

STEP 2: Using the notion of length for intervals in R, we construct *Lebesgue outer measure* which will be an outer measure (as will be defined in STEP 1). This will be defined for ALL subsets and should be viewed as the first attempt to construct Lebesgue measure. It will not be countably additive.

STEP 3: Show that the Lebesgue outer measure of an interval is its length.

STEP 4: (Caratheodory's Extension theorem). Given an outer measure $m^\star$ on an arbitrary set X, there is a σ-algebra $\mathcal{M}$ so that $m^\star$ restricted to $\mathcal{M}$ is a complete measure. (This statement as stated here is completely trivial since we could take $\mathcal{M}$ to be $\{\emptyset, X\}$; the proper version of this theorem will be stated later when we introduce some more concepts.)

STEP 5: Show that for Lebesgue outer measure on R, the $\mathcal{M}$ which will be constructed in Step 4 contains $\mathcal{B}$.

As we will show, these steps readily yield Theorem 3.20.

STEP 1: Define the concept of **outer measure**.

Definition 3.21. *An* **outer measure** *on a set X is a function $\mu^\star$ from $\mathcal{P}(X)$ to $[0,\infty]$ satisfying*
(i). $\mu^\star(\emptyset) = 0$.
(ii). $A \subseteq B$ implies that $\mu^\star(A) \le \mu^\star(B)$.
(iii). Given $A_1, A_2, \ldots$

$$\mu^\star(\bigcup_{n=1}^{\infty} A_n) \le \sum_{n=1}^{\infty} \mu^\star(A_n).$$

Remark: (ii) and (iii) are called, as before, monotonicity and subadditivity. $\mu^\star$ is **not** required to be countably additive, so $\mu^\star$ is **not** usually a measure on $\mathcal{P}(X)$.

STEP 2: Definition of *Lebesgue outer measure.* (If I is an interval, we let $|I|$ denote its length.)

Let $X = R$, $A \subseteq X$ and define

$$\mu^\star(A) := \inf\{\sum_{i=1}^{\infty} |I_i| : I_1, I_2, \ldots \text{ are open intervals with } A \subseteq \bigcup_i I_i\}.$$

Idea: You want to cover A by intervals as "efficiently as possible" making the sum of the lengths as small as possible.

$\mu^\star$ is called *Lebesgue outer measure* and our first task is to prove it is in fact an outer measure as defined in STEP 1.

Theorem 3.22. *$\mu^\star$ is an outer measure on R.*

Proof:
(i). This is trivial.
(ii). This is essentially trivial since any interval covering of B is an interval covering of A and hence in the definition of $\mu^\star(A)$, one is taking an infimum over a larger collection and hence the infimum is no larger.
(iii). (This takes a little more work.)

Case 1. $\mu^\star(A_n) = \infty$ for some n.
Then the inequality trivially holds.

Case 2. $\mu^\star(A_n) < \infty$ for all n.
Let $\epsilon > 0$. For each A_j, choose open intervals $I_1^j, I_2^j, I_3^j, \ldots$ so that $A_j \subseteq \bigcup_{i=1}^\infty I_i^j$ and

$$\sum_{i=1}^{\infty} |I_i^j| \leq \mu^\star(A_j) + \epsilon/2^j.$$

Now consider the countable collection of open intervals $\{I_i^j\}_{i,j\geq 1}$. Since the union of these contain each A_j, they contain $\bigcup_j A_j$. We therefore have

$$\mu^\star(\bigcup_j A_j) \leq \sum_{i,j=1} |I_i^j| = \sum_{j=1}(\sum_{i=1} |I_i^j|) \leq \sum_{j=1}(\mu^\star(A_j) + \epsilon/2^j) = \sum_{j=1} \mu^\star(A_j) + \epsilon.$$

Looking at the first and last term, since this inequality holds for all $\epsilon > 0$, we get

$$\mu^\star(\bigcup_j A_j) \leq \sum_{j=1} \mu^\star(A_j).$$

QED

STEP 3: Show that the Lebesgue outer measure of an interval agrees with its length.

Theorem 3.23. *For each finite interval I, we have*

$$\mu^\star(I) = |I|.$$

Proof:
It is enough to prove this for closed intervals $I = [a, b]$. (Why is this enough?)

$\leq$ is easy. For each ϵ, $[a, b] \subseteq (a - \epsilon, b + \epsilon)$ and hence $\mu^\star(I) \leq b - a + 2\epsilon$. Since this inequality is true for each ϵ, we get $\mu^\star(I) \leq b - a$.

The reverse inequality is a little harder since one needs to show that you *cannot* cover an interval $[a, b]$ by a union of intervals in some *tricky way* so that the sum of the lengths of these intervals is less than $b - a$.

Assume $[a, b] \subseteq \bigcup_i I_i$. By compactness we can find an integer N so that $[a, b] \subseteq \bigcup_{i=1}^N I_i$. To complete the proof we need to show that

$$b - a \leq \sum_{i=1}^{N} |I_i|$$

which is very *believable* to say the least. See the picture for the proof.
QED

STEP 4: Caratheodory's Theorem (This is the most difficult step.) Before we can even state this (recall that we mentioned earlier that the version given earlier was trivial and that we would have to modify it), we need to introduce some crucial definitions.
We assume here that we have a set X (not necessarily R) and an outer measure $\mu^\star$ on X.

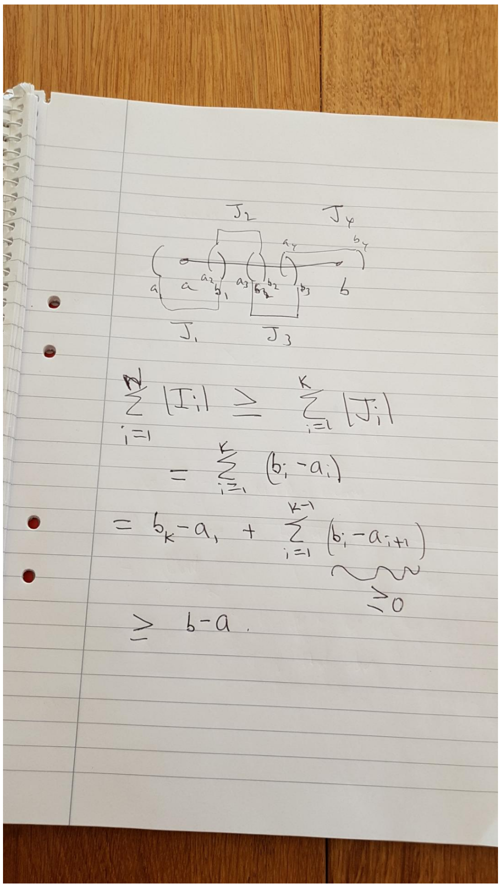
Definition 3.24. *If $\mu^\star$ is an outer measure on X, we call a subset $A \subseteq X$* **$\mu^\star$-measurable** *(see picture) if for all $E \subseteq X$,*

$$\mu^\star(E) = \mu^\star(E \cap A) + \mu^\star(E \cap A^c).$$

Remark: $\leq$ holds by subadditivity for all A and E. The reverse inequality holds trivially if $\mu^\star(E) = \infty$ and so we can assume that $\mu^\star(E)$ is finite.

Now we can state

Theorem 3.25. *(Caratheodory's Theorem) If $\mu^\star$ is an outer measure on X, then the collection $\mathcal{M}$ of $\mu^\star$-measurable sets is a σ-algebra and $\mu^\star$ restricted to $\mathcal{M}$ is a measure, which is also complete.*



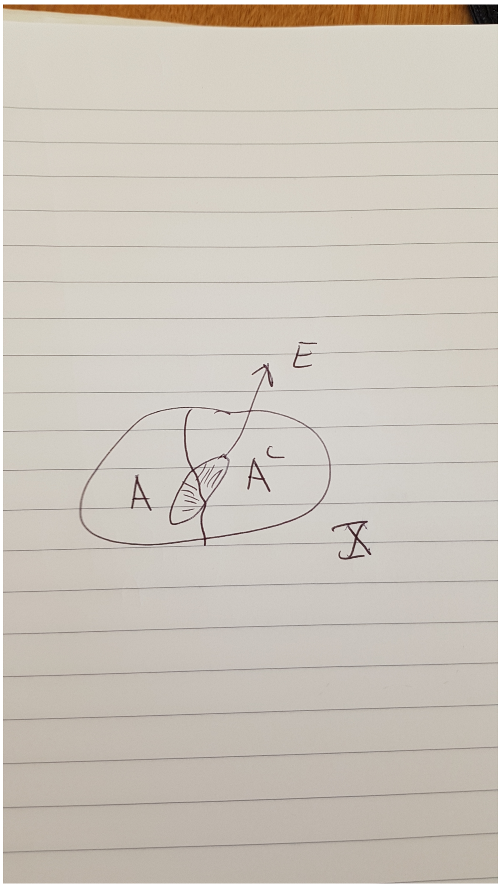
The proof of this takes some time and so it will be given its own subsection after this and right now we continue with the rest of the argument for constructing Lebesgue measure. It is also important to separate out Caratheodory's Theorem since then it will be available to us for use beyond just constructing Lebesgue measure.

STEP 5: Show that for Lebesgue outer measure on R, the $\mathcal{M}$ which was constructed in Step 4 contains $\mathcal{B}$.

So we need to show that each Borel set in R is $\mu^\star$-measurable (as defined in STEP 4) where $\mu^\star$ is the Lebesgue outer measure on R defined in STEP 2.

Since $\mathcal{M}$ is a σ-algebra (once we have proved STEP 4) and $\mathcal{B}$ is the smallest σ-algebra containing the open intervals, we just need to show that each open interval belongs to $\mathcal{M}$. To do that, it is enough to show each $(-\infty, a)$ and (b, ∞) are in $\mathcal{M}$ (since we can get any interval by intersecting two of these). We just do this for $(-\infty, a)$.

Let $E \subseteq R$ and we many assume $a \notin E$ as one point won't affect outer



measure. So we need to show

$$\mu^\star(E) \geq \mu^\star(E \cap (-\infty, a)) + \mu^\star(E \cap (a, \infty)). \tag{1}$$

Let $\{I_i\}$ be an arbitrary covering of E by open intervals. Let $I_i' := I_i \cap (-\infty, a)$ and $I_i'' := I_i \cap (a, \infty)$ and note that $\{I_i'\}$ ($\{I_i''\}$) is an interval covering of $E \cap (-\infty, a)$ ($E \cap (a, \infty)$) by open intervals. Hence we obtain

$$\sum_i |I_i| = \sum_i |I_i'| + |I_i''| = \sum_i |I_i'| + \sum_i |I_i''| \geq \mu^\star(E \cap (-\infty, a)) + \mu^\star(E \cap (a, \infty)).$$

Since the LHS is $\geq$ the RHS for all coverings of E by open intervals, we can take the infimum of the LHS over all such coverings and obtain (1).
QED

FINAL STEP: PUTTING IT ALL TOGETHER TO CONSTRUCT LEBESGUE MEASURE.
On R, we defined an outer measure $\mu^\star$ (Lebesgue outer measure) in STEP 2.

By STEP 4, we obtain a measure space $(R, \mathcal{M}, \mu^\star|_{\mathcal{M}})$ where $\mathcal{M}$ is the set of $\mu^\star$-measurable sets. By STEP 3, $\mu^\star(I) = |I|$ for all intervals I. By STEP 5, $\mathcal{B} \subseteq \mathcal{M}$. Hence we can restrict $\mu^\star$ from $\mathcal{M}$ down to $\mathcal{B}$ obtaining the desired measure space $(R, \mathcal{B}, \mu^\star|_{\mathcal{B}})$.

Finally, it is clear from the definition of the outer measure that $\mu^\star(A + x) = \mu^\star(A)$ for all sets A and $x \in R$. Hence $\mu^\star|_{\mathcal{B}}$ (as well as $\mu^\star|_{\mathcal{M}}$) is translation invariant.
QED

Remark: When someone refers to "Lebesgue measure on R", they may either be referring to $(R, \mathcal{B}, \mu^\star|_{\mathcal{B}})$ or to $(R, \mathcal{M}, \mu^\star|_{\mathcal{M}})$.

3.4 Proof of Caratheodory's Theorem

The proof of this is broken into a number of steps.

a. **$\mathcal{M}$ is an algebra.**
(i). $\emptyset \in \mathcal{M}$ is immediate.
(ii). $\mathcal{M}$ is closed under complementation since the definition is symmetric in A and A^c.
(iii). We need to show $A, B \in \mathcal{M}$ implies $A \cup B \in \mathcal{M}$.
Fix $E \subseteq X$. Noting that

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

and that this is a disjoint union, we have, using subadditivity,

$$\mu^\star(E \cap (A \cup B)) + \mu^\star(E \cap (A \cup B)^c) \leq$$

$$\mu^\star(E \cap (A \cap B)) + \mu^\star(E \cap (A^c \cap B)) + \mu^\star(E \cap (A \cap B^c)) + \mu^\star(E \cap (A^c \cap B^c)).$$

Using measurability of A applied to $E \cap B$ for the sum of the first two terms and applied to $E \cap B^c$ for the sum of the second two terms, this equals

$$\mu^\star(E \cap B) + \mu^\star(E \cap B^c) = \mu^\star(E)$$

where the last equality follows from the measurability of B. Hence $A \cup B \in \mathcal{M}$.

b. **$\mu^\star$ is finitely additive on $\mathcal{M}$.**
If $A, B \in \mathcal{M}$ are disjoint, then using measurability of A, we have

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

Now use induction. (Note that only one of the two sets was required to be measurable for this.)

c. $\mathcal{M}$ **is a** σ**-algebra .**
Since $\mathcal{M}$ is an algebra, it suffices (why??) to show that if $A_1, A_2, \ldots \in \mathcal{M}$ are disjoint, then $\bigcup_i A_i \in \mathcal{M}$. Now, let $B_n := \bigcup_{i=1}^n A_i$ and $B := \bigcup_i A_i$. We have, using measurability of A_n, that for all $E \subseteq X$,

$$\mu^\star(E\cap B_n) = \mu^\star(E\cap B_n\cap A_n)+\mu^\star(E\cap B_n\cap A_n^c) = \mu^\star(E\cap A_n)+\mu^\star(E\cap B_{n-1}).$$

This argument can be repeated inductively to obtain

$$\mu^\star(E \cap B_n) = \sum_{i=1}^n \mu^\star(E \cap A_i). \tag{2}$$

(Note that this is a slight variation of the finite additivity established in the previous step.)

Now, using measurability of B_n together with (2), we have that for any n

$$\mu^\star(E) = \mu^\star(E \cap B_n) + \mu^\star(E \cap B_n^c) = \sum_{i=1}^n \mu^\star(E \cap A_i) + \mu^\star(E \cap B_n^c) \geq$$

$$\sum_{i=1}^n \mu^\star(E \cap A_i) + \mu^\star(E \cap B^c).$$

Now looking at the left side and the right side and letting $n \to \infty$, we obtain

$$\mu^\star(E) \geq \sum_{i=1}^\infty \mu^\star(E \cap A_i) + \mu^\star(E \cap B^c) \geq \mu^\star(E \cap B) + \mu^\star(E \cap B^c) \tag{3}$$

where we used subadditivity and the definition of B in the last inequality. This establishes that $B \in \mathcal{M}$ and therefore that $\mathcal{M}$ is a σ-algebra .

d. $\mu^\star$ **is countably additive on** $\mathcal{M}$**; i.e.** $(X, \mathcal{M}, \mu^\star)$ **is a measure space.**
Let $A_1, A_2, \ldots \in \mathcal{M}$ be disjoint and let B_n and B be as defined in the previous step.

Note that by subadditivity, the last two terms in (3) is $\geq \mu^\star(E)$ and so we conclude we must have equalities everywhere. In particular, taking $E = B$, we obtain

$$\mu^\star(B) = \sum_i \mu^\star(A_i)$$

as desired.

e. **The measure space $(X, \mathcal{M}, \mu^\star)$ is complete.**
One first observes that any $A \subseteq X$ with $\mu^\star(A) = 0$ is $\mu^\star$-measurable since for any subset E

$$\mu^\star(E \cap A) + \mu^\star(E \cap A^c) = \mu^\star(E \cap A^c) \leq \mu^\star(E).$$

Hence if we have $B \in \mathcal{M}$, $\mu^\star(B) = 0$ and $A \subseteq B$, it follows that $\mu^\star(A) = 0$ and hence from the above $A \in \mathcal{M}$, as desired.
QED

Remark: Clearly any countable subset of R has Lebesgue measure 0. One might contemplate whether there are also uncountable subsets of R which have Lebesgue measure 0.

Historical remark: Lebesgue's original construction of Lebesgue measure was slightly more complicated that this approach. Caratheodory developed this approach some years after Lebesgue.

3.5 Relation between $\mathcal{B}$ and $\mathcal{M}$

For R, we know that $\mathcal{B} \subseteq \mathcal{M}$ and it is natural to ask if the two classes are the same. It turns out they are not. There are Lebesgue measurable sets which are not Borel sets. We won't prove this. There are different ways to do this: one can show that the two collection of sets have different cardinalities or one can show that $(R, \mathcal{B}, m)$ is not a complete measure space. Both take some work.

3.6 Uniqueness of Lebesgue measure on $\mathcal{B}$

We restrict to $[0, 1]$ for simplicity. We have constructed a measure on $\mathcal{B}_{[0,1]}$ which agrees with "length" on intervals. It is natural to ask if this measure is unique or whether there could exist a different measure on $([0, 1], \mathcal{B}_{[0,1]})$ which agrees with "length" on intervals. It turns out that it is unique and Dynkin's $\pi - \lambda$ Theorem will allow us to conclude this.

Theorem 3.26. *(Uniqueness of Lebesgue measure on the Borel sets)*
Let X be a set and $\mathcal{I}$ be a π-system on X. If μ_1 and μ_2 are two measures on $(X, \ \sigma(\mathcal{I})\ )$ such that

$$\mu_1(X) = \mu_2(X) < \infty$$

and

$$\mu_1(I) = \mu_2(I) \ \forall I \in \mathcal{I}$$

Then $\mu_1 = \mu_2$.

Applying this to $X = [0, 1]$ and $\mathcal{I}$ being the set of open intervals implies that there is only one measure on $([0, 1], \mathcal{B}_{[0,1]})$ which agrees with "length" on intervals.

Proof:
Assume μ_1 and μ_2 are two such measures. Let

$$\mathbf{D} := \{A \in \sigma(\mathcal{I}) : \mu_1(A) = \mu_2(A)\}.$$

Our goal is to show that $\mathbf{D} = \sigma(\mathcal{I})$. (Of course we have $\subseteq$.)

Step 1: $\mathbf{D}$ is a $\mathcal{D}$-system. (Proof at end.)

Step 2. Observe that $\mathcal{I} \subseteq \mathbf{D}$ by assumption.

Step 3. Using Dynkin's $\pi - \lambda$ Theorem for the equality and steps 1 and 2 for the containment below, we have

$$\sigma(\mathcal{I}) = \mathcal{D}(\mathcal{I}) \subseteq \mathbf{D}$$

and hence $\mu_1 = \mu_2$.

Lastly, we verify Step 1.
a. $X \in \mathbf{D}$ by assumption.
b. $A, B \in \mathbf{D}$ with $A \subseteq B$ implies that

$$\mu_1(B\backslash A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B\backslash A)$$

and hence $B\backslash A \in \mathbf{D}$.
c. If $E_1 \subseteq E_2 \subseteq E_3, \ldots$ and $E_i \in \mathbf{D}$ for all i, then using continuity from below for both measures, we have

$$\mu_1(\bigcup_i E_i) = \lim_{n\to\infty} \mu_1(E_n) = \lim_{n\to\infty} \mu_2(E_n) = \mu_2(\bigcup_i E_i)$$

and hence $\bigcup_i E_i \in \mathbf{D}$.
a,b, and c imply that $\mathbf{D}$ is a $\mathcal{D}$-system.
QED

One might guess that if one has two finite measures on X with the same total mass which agree on a collection of sets $\mathcal{E}$, then they agree on $\sigma(\mathcal{E})$.

Exercise: Find a counterexample to such a statement. It suffices to use an X with four elements.
(Hint: If you know some probability, one such example corresponds to the fact that a 2-dimensional random vector is not determined by its two 1-dimensional marginal distributions.)

This exercise illustrates the importance of the notion of a π-system.

3.7 Nonmeasurable sets

It is natural, for the case of Lebesgue outer measure, to ask if $\mathcal{M}$ contains all sets, that is, whether all sets are Lebesgue measurable.

Recall also our question at the start of these notes which asked if we could assign a "length" ℓ to all subsets of R satisfying
1. $\ell([a, b]) = b - a$.
2. If $A_1, A_2, \ldots$ are disjoint sets, then

$$\ell(\bigcup_i A_i) = \sum_i \ell(A_i)$$

3.

$$\ell(A + x) = \ell(A)$$

for all sets $A \subseteq R$ and $x \in R$.

These turn out to be closely related questions. We will prove the answer to the second question is no (assuming the axiom of choice). Note that this immediately implies that $\mathcal{M}$ cannot be everything since if it were, then $(R, \mathcal{M}, m)$ would be an example of a translation invariant measure on all subsets agreeing with length on intervals.

Theorem 3.27. *There does not exist a translation invariant measure on all subsets of R which gives length for intervals.*

Proof:
Assume μ is such a measure. Define an equivalence relation $\sim$ on $[0, 1]$ by

$$x \sim y \text{ if } x - y \in Q \text{ (Q denotes the rational numbers)}$$

Each equivalence class is countable and so the number of equivalence classes is uncountable. Let A consist of one element from each of the equivalence classes. (The Axiom of Choice allows us to contruct the set A; you can read about this axiom in Folland.) We now ask what must $\mu(A)$ be.

Step 1. $\mu(A) > 0$.
Subproof.
We claim that

$$[0, 1] \subseteq \bigcup_{q \in [-1,1] \cap Q} (A + q) \tag{4}$$

To see this, if $x \in [0, 1]$, choose $y \in A$ with $x - y \in Q$. Then $x = y + (x - y) \in A + (x - y)$. Since $x - y \in [-1, 1] \cap Q$, we obtain (4). Now $\mu(A + q) = \mu(A)$ for all q by the assumed translation invariance. Therefore, if $\mu(A) = 0$, then, by countable additivity, the measure of the RHS of (4) would be 0, a contradiction, since μ of the LHS is 1. Hence $\mu(A) > 0$.

Step 2. $\mu(A) = 0$.
Subproof.
Clearly

$$\bigcup_{q \in [0,1] \cap Q} (A + q) \subseteq [0, 2] \tag{5}$$

The sets arising in the union on the left hand side are disjoint since if some element u belonged to both $A + q_1$ and $A + q_2$, we would have $u = a_1 + q_1 = a_2 + q_2$ with $a_1, a_2 \in A$. Then $a_1 - a_2 (= q_2 - q_1) \in Q$ which implies $a_1 \sim a_2$ and hence $a_1 = a_2$. This then gives $q_1 = q_2$ also. Hence the sets are disjoint. Each of the sets on the left hand side has measure $\mu(A)$ and hence if $\mu(A) > 0$, then the LHS would have infinite measure, contradicting the RHS has measure 2. Hence $\mu(A) = 0$.

The two steps obviously give us a contradiction.
QED

We have used (uncountable) Axiom of Choice to show there cannot be a translation invariant measure on all subsets of R. What happens if we don't allow the axiom of choice? The following statement (taken from Wikipedia which I hereby credit) is a perhaps cryptic answer to this question.

Wikipedia quotation
In 1970, Robert M. Solovay constructed Solovay's model, which shows that it is consistent with standard set theory, excluding uncountable choice, that all subsets of the reals are measurable. However, Solovay's result depends on the existence of an inaccessible cardinal, whose existence and consistency cannot be proved within standard set theory.

If you are not interested in set theory, you are probably satisfied with this and do not want to dig any deeper.

3.8 Finitely additive measures: An interesting aside

I will just touch on this interesting topic briefly so that people are aware of it. I will simply mention a few facts to whet your appetite.

1. There exists a *finitely* additive "measure" on all subsets of R which is translation invariant and assigns length to intervals.

I put measure in quotes since it is only finitely additive and not countable additive. One can prove this by using some version of what is called the Hahn-Banach Theorem from functional analysis.

2. There exists a finitely additive "measure" on all subsets of R^2 which is invariant under all isometries of space and assigns standard volume to squares.

3. There does NOT exist a finitely additive "measure" on all subsets of R^3 which is invariant under all isometries of space and assigns standard volume to cubes.

Exercise: Show that the existence of the Banach-Tarski Paradox implies 3 above. (2 would then imply that there are no Banach-Tarski Paradoxes in the plane.)

Remark: There is something interesting and nontrivial going on here. It turns out that the group theoretic properties of the isometry group of R^3 is very different from the group theoretic properties of the isometry group of R^2 and this is what is lying behind the fact that one needs to go to 3 dimensions to obtain Banach-Tarski Paradoxes.

3.9 Other important constructions of measures

The following definition and theorem might seem a little abstract but it is very useful for constructing various measure spaces such as finite product spaces as well as infinite product spaces, the latter being needed for constructing an infinite number of independent random variables in probability theory.

Definition 3.28. *If X is a set and $\mathcal{A}$ is an algebra on X, a function μ_0 from $\mathcal{A}$ to $[0,\infty]$ is called a* **premeasure** *if*
1. $\mu_0(\emptyset) = 0$ and
2. If $A_1, A_2, \ldots,$ are (pairwise) disjoint elements of $\mathcal{A}$ **and** *$\bigcup_i A_i \in \mathcal{A}$, then*

$$\mu_0(\bigcup_i A_i) = \sum_i m(A_i).$$

Remark: If $\mathcal{A}$ were a σ-algebra , then this would just be a measure.

Theorem 3.29. *(Theorem 1.14 in F) If μ_0 is a premeasure on $(X, \mathcal{A})$, then there exists a measure μ on $(X, \sigma(\mathcal{A}))$ with $\mu(A) = \mu_0(A)$ for all $A \in \mathcal{A}$. If μ_0 is σ-finite on X, then μ is unique. (Uniqueness can fail in the non-σ-finite case.)*

Outline of the Proof:

1. Define an outer measure on all subsets of X by

$$\mu^\star(E) := \inf\{\sum_{i=1}^{\infty} \mu_0(A_i) : A_1, A_2, \ldots \in \mathcal{A},\ E \subseteq \bigcup_i A_i\}.$$

Verify, exactly as for Lebesgue outer measure, that this yields an outer measure.

2. Caratheodory's Theorem implies that $(X, \mathcal{M}, \mu)$ is a measure space where $\mathcal{M}$ is the σ-algebra of $\mu^\star$-measurable sets and $\mu := \mu^\star|_{\mathcal{M}}$.

3. Are we done with the proof? NO! We need to know that $\mathcal{A} \subseteq \mathcal{M}$ and that $\mu(A) = \mu_0(A)$ for $A \in \mathcal{A}$ and then we would be almost done. Did we have to do these two steps for the construction of Lebesgue measure? Yes! They were precisely STEPS 3 and 5. These two facts are proved in a very similar way here using the fact that we have a premeasure. Proposition 1.13 in Folland does exactly this. Finally we just restrict the measure obtained on $\mathcal{M}$ to $\sigma(\mathcal{A})$, completing the first statement.

For the second statement, if the measure space is finite, one can use the Dynkin's $\pi - \lambda$ Theorem to prove it using exactly the argument we gave for uniqueness of Lebesgue measure on the unit interval. Given the result for finite measure spaces, one can easily extend it to the σ-finite case by applying the above to each "finite piece". (The argument for this given in F is different.)
QED

3.10 Distribution functions for Borel measures on $[0, 1]$ and R; an application of the previous section

In this subsection, for simplicity, we stick to $[0, 1]$ (rather than R) and consider finite measures on $([0, 1], \mathcal{B}_{[0,1]})$ which are called Borel measures.

The following pretty easy proposition will start us off on this discussion.

Proposition 3.30. *Let μ be a finite Borel measure on $[0, 1]$ and define $F : [0, 1] \to [0, \mu([0, 1])]$ by*

$$F(x) := \mu([0, x]).$$

Then F is a weakly increasing and right continuous.

Proof:
Monotonicity of measures implies F is weakly increasing. The right continuity of F follows from continuity from above. For fixed t,

$$\lim_{s \downarrow t} F(s) = \lim_{s \downarrow t} \mu([0, s]) = \mu([0, t]) = F(t). \tag{6}$$

QED

Exercises:
1. Verify the middle equality in detail.
2. Show that we don't necessarily have left continuity of F and moreover that

$$F(t) - \lim_{s \uparrow t} F(s) = \mu(\{t\}). \tag{7}$$

The next result allows us to go the other way around.

Proposition 3.31. *Let F be a nonnegative weakly increasing and right continuous function on $[0,1]$ mapping into $[0,\infty)$. Then there exists a unique finite Borel measure μ on $[0,1]$ satisfying*

$$F(x) := \mu([0,x]).$$

Outline of Proof:
We consider the algebra $\mathcal{A}$ of subsets of $[0,1]$ consisting of a finite union of disjoint intervals which are open on the left and closed on the right and we also allow the set $\{0\}$. One checks that this is an algebra. Given a half open interval $I = (a,b]$, we let

$$\mu_0(I) := F(b) - F(a)$$

and we define μ_0 of a finite number of disjoint intervals just by adding up the above. (Technical messy point: you can represent a set by a union of disjoint such intervals in different ways and one needs to check that you always get the same number when you add these up. It's intuitive, easy to check but one should be aware that one needs to check it.) Also let $\mu_0(\{0\}) := F(0)$. Then one has to check that this is a premeasure on $\mathcal{A}$. Having done that, one can apply Theorem 3.29 to give us a unique measure μ on the Borel sets and then one checks that

$$F(x) := \mu([0,x]).$$

QED

Remark:
If $F(x) = x$, we get Lebesgue measure from this.

The bottom line of 2 which we should keep in mind is that there is a 1-1 correspondence between finite Borel measures on $[0,1]$ and functions F with the properties described above. Once we develop integration theory, this correspondence will basically define for us the so-called Lebesgue-Stieltjes integral.

Notation The F corresponding to a measure μ as above will be denoted by F_μ and the μ corresponding to a given F as above will be denoted by μ_F.

More remarks:
1. (6) and (7) imply that μ is continuous on $(0,1]$, meaning no atoms, if and only if F_μ is continuous. (Note that since F is only defined on $[0,1]$, an atom at 0 would not correspond to a discontinuity of F; e.g., $F \equiv 1$ corresponds to δ_0, a point mass at 0.)
2. If μ is a finite measure on $(R, \mathcal{B})$, one can define $F : R \to [0, \mu(R)]$ by $F(x) := \mu((-\infty, x])$ and one can go the other way in an analogous way.
3. Lebesgue measure on R does not satisfy the above; however, even for infinite measures, provided they are finite on compact sets, one can define an associated F, but now it is well defined only up to an additive constant. Therefore, for simplicity, we are sticking to $[0,1]$ with a finite measure.
4. If X is a "random variable" (whatever that means), its "law" or "distribution" is a measure on R with total measure 1 and the associated F is known as the "distribution function of X".

We end this subsection with a nice example to keep in mind. There are three types of distributions and we will explain later on that every distribution is a combination (more precisely a convex combination) of three distributions of these types. The first type is called absolutely continuous (this will be defined later) and is exemplified by Lebesgue measure corresponding $F(x) = x$. From a probability point of view, these correspond to having a probability density function. The second type are called the atomic distributions which is exemplified by the example right below. The third (and most subtle) type are called continuous singular and will exemplified by the Cantor ternary function in the next subsection.

We now give one example of an atomic measure. Order the rational numbers in $[0,1]$, $q_1, q_2, \ldots$ and we want to put a weight of $1/2^n$ on q_n. This yields the measure μ which satisfies

$$\mu(A) := \sum_{i: q_i \in A} 1/2^i$$

for all Borel sets A (check this).

Exercises:
1. Check that the corresponding distribution function F satisfies

$$F(x) = \sum_{i: q_i \leq x} 1/2^i.$$

2. Check that this is right continuous at all points and is discontinuous from the left at exactly the rational numbers.
3. Observe or check that this measure is atomic (as defined earlier in the chapter) since $\mu(Q^c) = 0$ and the atoms are exactly Q.

Remark: It will be a consequence of Lebesgue's Theorem (much later on) that F is differentiable Lebesgue-a.e. However it is not the case that we have differentiability at all of the irrational numbers.

3.11 An overview of the Cantor set, the Cantor ternary function and the "Cantor measure"

The last term is put in quotes since I don't think that is a standard expression.

We will now see the previous section (where we associate Borel measures on $[0, 1]$ with weakly increasing right continuous functions) come alive by looking at a very specific but enlightening example. This section will be an overview and details can be obtain in F. It is also a teaser for what will be coming later on.

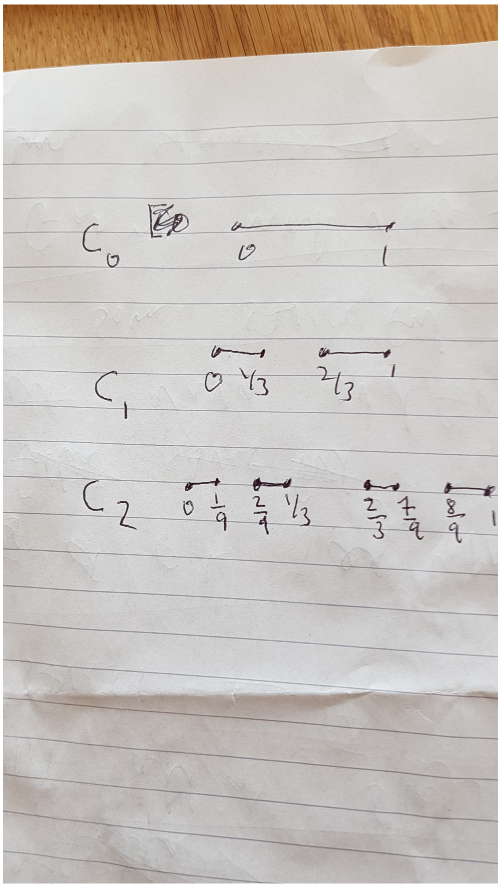
We first tell/remind the reader what the Cantor set is; it is constructed iteratively.

Let $C_0 = [0, 1]$.
Let C_1 be C_0 with the middle third removed $(= [0, 1]\backslash(1/3, 2/3))$.
Let C_2 be obtained from C_1 by removing the middle third of each interval.
One continues defining $C_3, \ldots$. (See picture). Note C_n consists of 2^n disjoint closed intervals each of length $1/3^n$.

Definition 3.32. *The Cantor set, C, is defined to be $\bigcap_n C_n$.*

The following proposition gives the main important properties of C.

Proposition 3.33. *1. C is a nonempty compact set.*
2. The Lebesgue measure of C is 0.
3. It has no isolated points. (If A is set, $x \in A$ is called isolated *if for some $\epsilon > 0$, $A \cap (x - \epsilon, x + \epsilon) = \{x\}$.)*
4. 1 and 3 together mply that C is uncountable.
5. C is also the set of points in $[0, 1]$ whose ternary expansion (which consists of 0's, 1's and 2's) only contains 0's and 2's. (When the ternary expansion is not unique, eg. 1/3, we require that only one *of the two expansions has this property, which would then be the nonterminating one.)*



Proof outline of some parts:
1. C_n is closed and hence, by elementary topology, C is a nonempty compact set.
2. From the observation earlier, C_n has Lebesgue measure $(2/3)^n$ and hence C has measure 0.
3. and 5. skip.
QED

Exercise: First note that C cannot contain an interval since its Lebesgue measure is 0. Interestingly, there also exist closed sets with *positive* Lebesgue measure which do not contain an interval. Construct one by modifying the construction of the Cantor set by removing "centered" intervals of length smaller than 1/3. Such sets are called "fat Cantor sets". Another construction, perhaps simpler, is simply to remove smaller and smaller intervals around the rational numbers.

There is a natural measure of total weight 1 on C. It gives measure $(1/2)^n$ to each of the 2^n intervals of length $1/3^n$. Note each of these "level-n" intervals are given a measure which is much more than its length. They are given a measure which is $(3/2)^n$ times its length. One can with some work construct this measure μ_C by defining it as above on our "basic intervals" and extending it to all Borel sets.

The important feature of this measure is that it will have no atoms and it will give all of its weight to C (meaning no weight to C^c), a set of Lebesgue measure 0. Such measures are called **continuous singular** (a concept we will come back to later on) and it might be surprising that they exist if you have not seen them before.

The Cantor Ternary function is then the distribution function, as defined in the previous subsection, of μ_C. This function, which we call F_C, has the fascinating properties that
(i). F is a weakly increasing function on $[0, 1]$ with $F(0) = 0$ and $F(1) = 1$.
(ii). F is continuous.
(iii). $F' = 0$ (Lebesgue)-a.e. on $[0, 1]$. (Why?)

Remarks:
a. Somehow the values of F manage to go from 0 up to 1 continuously as we move along $[0, 1]$ even though the derivative is 0 Lebesgue-a.e.
b. Note that the fundamental theorem of calculus

$$\int_0^1 F'dx = F(1) - F(0)$$

fails here! This failure of the fundamental theorem of calculus will be put into a more general context later on but we wanted to introduce this example already here.
c. The behavior of F' on the Cantor set itself is discussed in research papers.

3.11.1 The "dimension" of the Cantor set: An interesting Aside

It is often said that the Cantor set has "dimension" $\frac{\log 2}{\log 3}$. To state this precisely, one can introduce the precise notion of Hausdorff dimension (or of Minkowski dimension) and prove that these give $\frac{\log 2}{\log 3}$ for the Cantor set. However, we don't want to introduce these concepts in these notes. Rather, I will convince you that it is reasonable to declare that the dimension of the Cantor set is $\frac{\log 2}{\log 3}$ without giving any precise definitions.

The first key observation is that if we take a cube in R^d and scale it by the integer k, the number of copies of the cube we get is k^d . If want to recover d (the dimension), we can say that if we scale the cube by k and let $N = N(k)$ be the number of copies of the cube we get, then $d = \log(N)/\log(k)$ (and the

d didn't really depend on k). If we now look at the Cantor and scale it by 3 (so $k = 3$), what do we get? After a moment of thought, $3C$ consists of 2 copies of C, one being the original C and the other simply being C translated to the right by 2 and so sitting inside $[2, 3]$. There is nothing in $(1, 2)$ since the Cantor set doesn't intersect $(1/3, 2/3)$. So, by scaling by $k = 3$, we get $N = 2$ copies and so the dimension d "should be", using the above formula, $\frac{\log 2}{\log 3}$.

Bottom line: Just because a set has Lebesgue measure 0 does not mean that it is "trivial" and of no interest to us. Quite to the contrary, it can still have very interesting structure and there is much research studying the dimensions (and other properties) of various sets all of which might have Lebesgue measure 0.

3.12 The Borel-Cantelli Lemma

We almost end this chapter with the Borel-Cantelli Lemma, which is useful in many different situations, not the least of which is probability theory.

Definition 3.34. *If $E_1, E_2, \ldots$ is a sequence of measurable sets in a measure space, we let*

$$\limsup E_i := \bigcap_{n=1}^{\infty} (\bigcup_{k=n}^{\infty} E_k)$$

which is also often written as (E_n i.o.) with i.o. meaning **infinitely often** *since it means that x is contained inside of infinitely many E_n's.*

Exercise: If we consider instead

$$\liminf E_i := \bigcup_{n=1}^{\infty} (\bigcap_{k=n}^{\infty} E_k),$$

how does this differ from the above? Which is larger? What is an example where they differ?

Lemma 3.35. *((First) Borel-Cantelli Lemma)*
Let $E_1, E_2, \ldots$ be a sequence of measurable sets in the measure space $(X, \mathcal{M}, m)$. If $\sum_i m(E_i) < \infty$, then

$$m(\limsup E_i) = 0.$$

Proof:
For each n, we have by subadditivity that

$$m(\limsup E_i) \leq m(\bigcup_{k=n}^{\infty} E_i)) \leq \sum_{k=n}^{\infty} m(E_i).$$

Since this holds for each n and the RHS is the tail of a convergent series, we have that $m(\limsup E_i) = 0$.
QED

Remark: The following example shows that $\sum_i m(E_i) < \infty$ is not necessary for $m(\limsup E_i) = 0$. Consider the unit interval with Lebesgue measure and let $E_n := [0, 1/n]$. (Check this!)

3.13 An application of the Borel-Cantelli Lemma to diophantine approximation: An interesting Aside

Since the rationals Q are dense in R, we can approximate any number as well as we want by elements of Q. To make the question of approximation more quantitative (and more interesting!), we can ask instead whether we can approximate a real number x by a rational p/q where the approximation is small relative to the denominator q or some function of q. Here is a our first nontrivial result of this kind.

Theorem 3.36. *For every $\delta > 0$, the set*

$$A_\delta := \{x \in R : |x - \frac{p}{q}| \leq \frac{1}{q^{2+\delta}} \text{ for infinitely many positive integer pairs } (p, q)\}$$

has Lebesgue measure 0. (In words, almost no real number can be "well approximated by rationals" if "well approximated by rationals" means that there are infinitely many positive integer pairs (p, q) with $|x - \frac{p}{q}| \leq \frac{1}{q^{2+\delta}}$.)

Proof: It is enough to show that $m(A_\delta \cap [0, 1]) = 0$ since the same argument shows we can replace $[0, 1]$ by $[n, n + 1]$ and then we can use countable additivity. This will now follow very easily from the Borel-Cantelli Lemma. For positive integer pairs (p, q), we let

$$E_{p,q} := \{x \in [0, 1] : |x - \frac{p}{q}| \leq \frac{1}{q^{2+\delta}}\}$$

and note that clearly $m(E_{p,q}) \leq \frac{2}{q^{2+\delta}}$. By the Borel-Cantelli Lemma, it suffices to show that

$$\sum_{p,q} m(E_{p,q}) < \infty.$$

Note that for each fixed q, $E_{p,q}$ is nonempty for at most, say, $4q+1$ values of p since certainly p must belong to $[-2q, 2q]$ for $E_{p,q}$ to be nonempty. This gives, as needed,

$$\sum_{p,q} m(E_{p,q}) \leq \sum_{q} \frac{4q+1}{q^{2+\delta}} < \infty.$$

QED

One can either stop reading this subsection now or continue and read more about this interesting story.

Note that if $\delta_1 < \delta_2$, $A_{\delta_2} \subseteq A_{\delta_1}$ so that the statement becomes stronger as δ goes to 0. The following result complements the above result. We do not prove it but mention that it is fairly elementary and uses the pigeon hole principle from combinatorics.

Theorem 3.37. *Every irrational number belongs to*

$$A_0 := \{x \in R : |x - \frac{p}{q}| \leq \frac{1}{q^2} \text{ for infinitely many integer pairs } (p,q)\}$$

and so in particular a.e. x belongs to A_0. (In words, almost every real number can be "well approximated by rationals" if "well approximated by rationals" now means that there are infinitely many integer pairs (p,q) with $|x - \frac{p}{q}| \leq \frac{1}{q^2}$.)

Interestingly, the story does not end here. It would seem that nothing grows faster than q^2 but slower than $q^{2+\delta}$ for every $\delta > 0$ and so it seems we now know everything that needs to be know. But this is not true. $q^2(\log q)^\ell$ grows between these two rates and one can ask whether

$$A_{0,\ell} := \{x \in R : |x - \frac{p}{q}| \leq \frac{1}{q^2(\log q)^\ell} \text{ for infinitely many integer pairs } (p,q)\}$$

has 0 measure or full measure (meaning the complement has measure 0). Keep in mind that for any $\ell_1 < \ell_2$ and any $\delta > 0$, we have

$$A_\delta \subseteq A_{0,\ell_2} \subseteq A_{0,\ell_1} \subseteq A_0$$

and so the above results do not answer this question. It turns out that $A_{0,\ell}$ has measure 0 if if $\ell > 1$ and full measure if $\ell \leq 1$. This is a quite a refined picture but it can be made more refined. If we now replace $q^2(\log q)^\ell$ by $q^2(\log q)(\log\log q)^\ell$, again the answer to the question depends on whether $\ell > 1$ or $\ell \leq 1$. And one can keep going getting more and more refined results. What is really going on here? It turns out there is an "integrability condition" going on here which answers everything. The following is the definitive result and includes the above results.

Theorem 3.38. *Let $f := f(q)$ be an increasing sequence and consider*

$$A_f := \{x \in R : |x - \frac{p}{q}| \leq \frac{1}{q^2 f(q)} \text{ for infinitely many integer pairs } (p,q)\}$$

Then A_f either has measure 0 or full measure and these correspond exactly to the cases

$$\sum_{q=1}^{\infty} \frac{1}{qf(q)} < \infty \text{ or } = \infty.$$

The proof for the summable case is carried out exactly as was done for the proof of Theorem 3.36. The other direction is more difficult and we don't discuss.

Is that the whole story now? Not really! All the sets A_δ for $\delta > 0$ have measure 0 but as we mentioned in the section on the Cantor set, measure 0 sets can still be worthy of investigation. In particular, one can still ask how "large" A_δ is (in some sense) even though the Lebesgue measure of it is 0. Our last result tells us what the "fractal size" of A_δ is and more precisely what its Hausdorff dimension is. Although we are not defining the Hausdorff dimension of a set, it has a precise definition which one can google.

Theorem 3.39. *(Jarnik, Besicovitch) For each $\delta > 0$, the Hausdorff dimension of A_δ is given by $\frac{2}{2+\delta}$.*

3.14 Baire Category and topologically big: An interesting aside

If we have a measure space with total measure 1 and if we call a set *large* if its measure is 1, then clearly a countable intersection of large sets is large. In general, it seems reasonable to call a property of sets "large" if whenever we intersect a countable number of large sets, we get something which is large.

It turns out that there is a useful notion of "large" for complete metric spaces. One calls a set "large" if it contains a dense G_δ where we recall that a G_δ is, by definition, a countable intersection of open sets. For this to be a reasonable definition of being large, we should have that a countable intersection of such sets is itself one of these sets. This is true and nontrivial and is called the Baire Category Theorem which we will not prove. (We state it in a slightly stronger way than it is usually stated; however this version follows immediately from the usual version.)

Theorem 3.40. *Let X be a complete metric spaces and for each n, let U_n be a dense G_δ. Then*

$$\bigcap_{n=1}^{\infty} U_n$$

is a dense G_δ. (The fact it is a G_δ is trivial; the content is that it is dense.)

Being a complete metric space is crucial for this theorem. To see this, let $X := Q = \{q_1, q_2, \ldots\}$ be the rational numbers in $[0,1]$ and let $A_n := Q \backslash \{q_n\}$. Clearly each A_n is a dense G_δ (in fact dense and open). However $\bigcap_n U_n = \emptyset$ which is certainly not dense.

So we now have two different notions of "large" for subsets of $[0,1]$, namely having Lebesgue measure 1 and containing a dense G_δ. It turns out that these two notions are completely different. (The expression "large in the category sense" is sometimes used for the latter: this is not the same "category" in "category theory".)

Theorem 3.41. *There exists a set $A \subseteq [0,1]$ which is a dense G_δ in $[0,1]$ of Lebesgue measure 0.*

Remark: So A is large in the topological sense and small in the measure theoretic sense while its complement will be small in the topological sense and large in the measure theoretic sense.

Proof:
Let $Q = \{q_1, q_2, \ldots\}$ and let $U_n := \bigcup_{i=1}^{\infty}(q_i - \frac{1}{n2^i}, q_i + \frac{1}{n2^i})$. Note that each U_n is a dense open set and hence (by the Baire Category Theorem) $\bigcap_n U_n$ is a a dense G_δ. What is its Lebesgue measure? It is immediate that for each n, $m(U_n) \leq \frac{2}{n}$ implying that $\bigcap_n U_n$ has zero Lebesgue measure.
QED

Oxtoby has written a beautiful book "Category and Measure" which studies at length the relationship between these two notions of largeness.

Another nice example where these two different notions of being large differ is in diophantine approximation which we just discussed in the previous section. It turns out that the sets A_δ ($\delta > 0$) defined in Theorem 3.36, while having measure zero, contain in fact a dense G_δ. Hence although they are measure theoretically small, they are topologically large.

One calls an irrational number x *Liouville* if for all n, there are integers p and q such that

$$|x - \frac{p}{q}| < \frac{1}{q^n}.$$

It is easy to see that the set of Liouville numbers is contained in each A_δ and hence has measure 0 (and even Hausdorff dimension 0). (In fact, the Liouville numbers is just $\bigcap_\delta A_\delta$.)

On the other hand, the Liouville numbers can be expressed as

$$([0,1]\backslash Q) \cap \bigcap_n \bigcup_{p,q} (\frac{p}{q} - \frac{1}{q^n}, \frac{p}{q} + \frac{1}{q^n}).$$

This is clearly a dense G_δ.

3.15 Additive vs. Linear functions: An interesting Aside; this subsection actually requires the notion of a measurable function from the next chapter

We end this section with a fun application of exercise 31 in Chapter 2 of F. As we know, $f : R \to R$ is linear if

1. $f(x+y) = f(x) + f(y)$ for all $x, y \in R$

and

2. $f(cx) = cf(x)$ for all $x, c \in R$

In fact, in 1-d, it is easy to check that (2) implies (1) (even if you assume (2) for only one nonzero x). The question here is whether (1) implies (2). We call f additive if it satisfies (1) and so this question is whether being additive implies being linear. It turns out the answer is no but that if you make some very weak assumptions on f, then the answer becomes yes. We are more concerned with the latter here but we first show (1) does not imply (2). One needs to know something about abstract vector spaces.

Proposition 3.42. *There exists $f : R \to R$ which is additive but not linear.*

Proof outline:
View R as a vector space over Q (which is then infinite dimensional). One checks that the set $\{1, \sqrt{2}\}$ is linearly independent. (This follows easily from the fact that $\sqrt{2}$ is irrational.) Extend this set to a basis $\mathcal{B}$ for R (as a Q-vector space). Let f be a function from $\mathcal{B}$ to R which takes 1 to 1, $\sqrt{2}$ to 3 and is arbitrary on the other elements of $\mathcal{B}$. By linear algebra, f can be extended to a Q-linear transformation from R to R, meaning (1) holds and (2) holds for $c \in Q$. In particular, f is additive. But f cannot be R-linear,

since any such map $x \to ax$ which takes 1 to 1 is the identity.
QED
Remark:
The Axiom of choice is used here to construct the basis $\mathcal{B}$.

Our main theorem here is the following.

Theorem 3.43. *If $f : R \to R$ is additive and measurable, then f is linear.*

Before we prove this, we begin with a warm up which is weaker.

Proposition 3.44. *If $f : R \to R$ is additive and continuous, then f is linear.*

Proof:
One first observes that (1) implies (2) for rational c. To see this, (1) implies by induction that for positive integer n, $f(nx) = nf(x)$. Therefore if n, m are positive integers, then

$$f(m\frac{n}{m}x) = mf(\frac{n}{m}x)$$

giving

$$f(\frac{n}{m}x) = \frac{1}{m}f(nx) = \frac{n}{m}f(x).$$

(1) easily gives (why?) $f(-x) = -f(x)$ for all x and from this it is easy to see that (2) is true for all rational c, positive or negative.
It therefore follows that $f(x) = xf(1)$ for all rational x. Since Q is dense and the two sides are continuous in x (the first by assumption), they must be equal for all x. Hence f is linear.
QED

To prove Theorem 3.43, we begin with a lemma.

Lemma 3.45. *If f is additive but not linear, then f is unbounded in every interval about* 0*.*

Proof:
By scaling f and multiplying by -1 if necessary, we can assume that $f(1) = 1$. Fix the interval $(-\epsilon, \epsilon)$ around 0. If f is not linear, then there must exist x such that $f(x) \neq x$. We have seen above that (1) implies (2) for rational c and hence x must be irrational. Without loss of generality $x > 0$ (otherwise, one does a similar argument). We now use (without proof) an elementary

fact (called Kronecker's Theorem): namely, there exist integers n and m arbitrarily large so that

$$|nx - m| < \epsilon.$$

(Note that this is certainly implied by Theorem 3.37.) We then have

$$f(nx - m) = nf(x) - m = nx - m + n(f(x) - x).$$

If n, m are very large and $|nx - m| < \epsilon$, then, since $f(x) - x \neq 0$, we have a point in $(-\epsilon, \epsilon)$ whose f value becomes in absolute value as large as we want. QED

Proof of Theorem 3.43:
By the previous lemma, it suffices to show that f is bounded in some interval $(-\epsilon, \epsilon)$ around 0. Let $A_n = \{x : |f(x)| \leq n\}$. By our measurability assumption, the A_n's are measurable, clearly increasing and their union is R. By continuity of measure from below, $m(A_n) > 0$ for some n. Fixing such an n, by exercise 31 of chapter 2 in F, $(-\epsilon, \epsilon) \subseteq A_n - A_n$ for some $\epsilon > 0$. Then for all $x \in (-\epsilon, \epsilon)$, $x = a - b$ with $a, b \in A_n$ implying that

$$|f(x)| = |f(a - b)| = |f(a) - f(b)| \leq |f(a)| + |f(b)| \leq 2n.$$

QED

3.16 The Vitali–Hahn–Saks Theorem: An interesting Aside

Question: Let μ_n be a sequence of probability measures on $(X, \mathcal{M})$ and assume that for each $E \in \mathcal{M}$, the limit

$$\lim_{n \to \infty} \mu_n(E)$$

exists which we denote by $\mu_\infty(E)$. Is it the case that μ_∞ is a measure?

Exercises: (a). Show that the set function μ_∞ is finitely additive and that it gives measure 0 and 1 to $\emptyset$ and X.
(b). Try to prove μ_∞ is countably additive (i.e. is a measure) in order to understand what the difficulty in proving this is.

It turns out that the answer to the above question is yes as proved independently by the three above authors. The proof is nontrivial. One of the key tools is the Baire Category Theorem from subsection 3.14. The proof will be given in subsection 7.4 after we have more "meat on the bones".

4 The Lebesgue Integral and Integration Theory

4.1 A little on Riemann integration and the loose idea of Lebesgue integration

The Lebesgue integral, while it takes some time to define, is based on a very simple idea. This idea is to make a simple but essential modification to the construction of the Riemann integral.

Little reminder about the Riemann integral

I won't define or remind you of Riemann integration (look it up) but in a nutshell, the Riemann integral of a function f defined on $[0, 1]$ is obtained by partitioning the domain $[0, 1]$ into very small intervals, constructing an approximating Riemann sum and hoping that the limit exists as the length of these small intervals goes to 0, no matter how you take your Riemann sum. If the limit exists, then the function is called Riemann integrable (RI). The **crucial** fact is that one is breaking up the x-axis.

In calculus courses, it is stated without proof that continuous functions on $[0, 1]$ are RI. Also RI functions are necessarily bounded.

It turns out that if f is a bounded function on $[0, 1]$, then there is an interesting necessary and sufficient condition for f to be RI which was proved by Lebesgue.

Theorem 4.1. *If f is a bounded function, then f is RI if and only if the set $\{x : f \text{ is not continuous at } x\}$ has Lebesgue measure 0.*

Exercise: Show that for any function f on $[0, 1]$, $\{x : f \text{ is continuous at } x\}$ is a Borel set in $[0, 1]$.

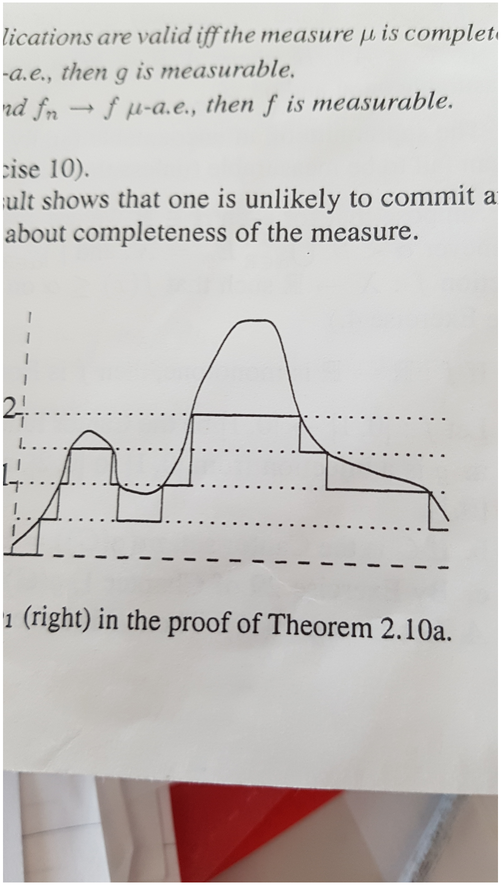
Remarks:
1. The classic example of a function which is not RI is I_Q, which is 1 on the rationals and 0 on the irrationals. (Of course, this is discontinuous at every point.) It will turn out that I_Q will be trivially Lebesgue integrable.
2. Interestingly, there is a function on $[0, 1]$ which is continuous at precisely the irrational numbers but there is no function on $[0, 1]$ which is continuous at precisely the rational numbers.

Idea of the Lebesgue integral on $[0, 1]$
Instead of breaking up the x-axis, we break up the y-axis. Let's say that f takes values in $[0, 1]$. We partition the y-axis $[0, 1]$ into $0 = a_0 < a_1 <$

$a_2 < \ldots < a_n = 1$ and approximate (see picture or draw some pictures for yourself) "the integral" by

$$\sum_{i=0}^{n-1} a_i m(\{x : f(x) \in [a_i, a_{i+1})\})$$

where m is Lebesgue measure (and we take the final interval to be closed on the right). If the maximum increment $\max_i\{a_{i+1} - a_i\}$ is small, we should get a good approximation of the "the integral".



Remark: When you break up the y-axis, then the structure of the domain is irrelevant which is why we can do this kind of integration on any measure space.

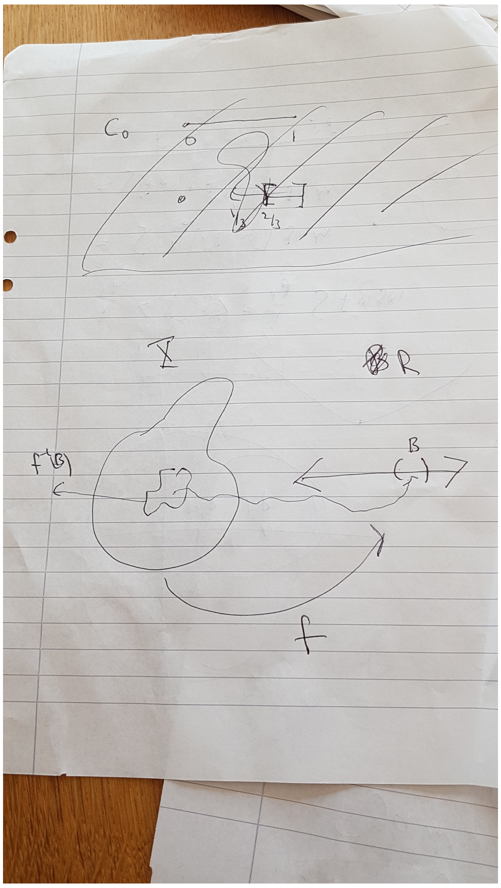
4.2 Measurable functions

We now introduce the class of functions that we will consider and which we will eventually assign an integral to. These are called "measurable" functions.

Exactly as you have seen that "almost all" sets are measurable, it is also the case that "almost all" functions are measurable (almost all here is not being used in a technical sense.)

Definition 4.2. *If* $(X, \mathcal{M})$ *is a measurable space, a mapping* $f : X \to R$ *is called* **measurable** *if for all* $B \in \mathcal{B}$ *(recall that* $\mathcal{B}$ *is the collection of Borel sets in* R*), we have that (see picture)*

$$f^{-1}(B) := \{x \in X : f(x) \in B\} \in \mathcal{M}.$$



Definition 4.3. *More generally, if* $(X, \mathcal{M})$ *and* $(Y, \mathcal{N})$ *are measurable spaces, a mapping* $f : X \to Y$ *is called* **measurable** *if for all* $E \in \mathcal{N}$

$$f^{-1}(E) \in \mathcal{M}.$$

The next proposition demonstrates that to check measurability, one does not always have to check the above property for all $E \in \mathcal{N}$ but just some suitable subclass.

Proposition 4.4. *Assume we have measurable spaces* $(X, \mathcal{M})$ *and* $(Y, \mathcal{N})$ *and a mapping* $f : X \to Y$. *Assume also that* $\mathcal{E} \subseteq \mathcal{N}$ *is some collection of sets which generates* $\mathcal{N}$ *(i.e.* $\sigma(\mathcal{E}) = \mathcal{N}$*). Then* $f : X \to Y$ *is measurable if for all* $E \in \mathcal{E}$

$$f^{-1}(E) \in \mathcal{M}.$$

(The proof is not so hard but the proof technique is very important to understand as it is used quite often.)

Proof:
Let

$$\mathcal{F} := \{E \in \mathcal{N} : f^{-1}(E) \in \mathcal{M}\}$$

Our goal is to show that $\mathcal{F} = \mathcal{N}$ and by assumption, $\mathcal{E} \subseteq \mathcal{F}$. If we can show that $\mathcal{F}$ is a σ-algebra , then it will follow that $\mathcal{F} \supseteq \sigma(\mathcal{E}) = \mathcal{N}$ and we would be done.

The fact that $\mathcal{F}$ is a σ-algebra follows from easy set theory together with the fact that $\mathcal{M}$ is a σ-algebra . $X, \emptyset \in \mathcal{F}$ is left to the reader.

Next, we have

$$E \in \mathcal{F} \text{ implies } f^{-1}(E) \in \mathcal{M} \text{ implies } (f^{-1}(E))^c \in \mathcal{M} \text{ implies } f^{-1}(E^c) \in \mathcal{M}$$

implies $E^c \in \mathcal{F}$ noting that $(f^{-1}(E))^c = f^{-1}(E^c)$ (Check this!).

Finally, we have

$$E_1, E_2, \ldots \in \mathcal{F} \text{ implies } f^{-1}(E_1), f^{-1}(E_2), \ldots \in \mathcal{M} \text{ implies } \bigcup_i (f^{-1}(E_i)) \in \mathcal{M}$$

$$\text{implies } f^{-1}(\bigcup_i E_i) \in \mathcal{M} \text{ implies } \bigcup_i E_i \in \mathcal{F}$$

noting that $\bigcup_i (f^{-1}(E_i)) = f^{-1}(\bigcup_i E_i)$ (Check this!).
QED

Remark: f^{-1} behaves much nicer than f in the sense that f^{-1} behaves well with respect to set theoretic operations while f does not.

Exercise:
1. Show $f^{-1}(A) \cap f^{-1}(B) = f^{-1}(A \cap B)$.
2. How do $f(A) \cap f(B)$ and $f(A \cap B)$ relate to each other?
3. How do $(f(A))^c$ and $(f(A^c))$ relate to each other? How about if we assume f is surjective?

Corollary 4.5. *A function $f : (X, \mathcal{M}) \to R$ is measurable if $f^{-1}(U) \in \mathcal{M}$ for all half infinite open intervals U. Note that this implies that a continuous function from R to R is Borel-measurable; i.e. $f : (R, \mathcal{B}) \to R$ is measurable.*

Since we will now only consider measurable functions, one would hope that when we apply simple operations to measurable functions, the result is still a measurable function. The following is a first step in this direction.

Proposition 4.6. *If $f, g : (X, \mathcal{M}) :\to R$ are measurable, then $f + g$ and fg are measurable.*

Proof:
For all $a \in R$, we have

$$\{x : (f+g)(x) > a\} = \bigcup_{q \in Q} (\{x : f(x) > q\} \cap \{x : g(x) > a - q\}) . \qquad (8)$$

$\supseteq$ is trivial. To see the opposite containment, if $x \in$ LHS, choose $q \in Q$ so that

$$0 < f(x) - q < f(x) + g(x) - a.$$

Then for this q, x will be in the corresponding set on the RHS establishing (8). Now, f, g being measurable implies each of the terms in the union are in $\mathcal{M}$ and since we have a *countable* union, the RHS and hence the LHS belong to $\mathcal{M}$. Hence $f + g$ is measurable.

For the second statement, one first observes that

$$fg = 1/2[(f+g)^2 - f^2 - g^2].$$

In view of the first part (and some trivial things like the negative of a measurable function is measurable and multiplying by a constant also results in a measurable function), it suffices to show that if f is measurable, then f^2 is measurable. To verify this, we note that

$$\{x : f^2(x) \geq c\} = X \text{ if } c \leq 0$$

and

$$\{x : f^2(x) \geq c\} = \{x : f(x) \geq c^{1/2}\} \cup \{x : f(x) \leq -c^{1/2}\} \text{ if } c > 0.$$

QED

Remark: A key feature of the first part of the above proof is that we always have to express things in terms of *countable* operations since these are the only things allowed in measure theory.

It sometimes becomes natural (among other reasons since we are taking limits of various things) to consider mappings from $(X, \mathcal{M})$ to $\overline{R} := R \cup \{-\infty, \infty\}$. Measurability is defined as before: for all $c \in R$,

$$\{x \in X : f(x) > c\} \in \mathcal{M}.$$

Exercise: Show that if $f : (X, \mathcal{M}) \to \overline{R}$ is measurable, then $\{x \in X : f(x) = \infty\} \in \mathcal{M}$.

Limiting procedures are going to be central to this whole theory; hence we would want that limits of measurable functions are measurable. We will see this fact soon.

Proposition 4.7. *If $f_1, f_2, \dots$ is a sequence of measurable functions, then $\sup_j f_j$ is a measurable function where $\sup_j f_j$ is defined (being a bit pedantic) by*

$$(\sup_j f_j)(x) := \sup_j (f_j(x)).$$

The same result holds for $\inf_j f_j$ defined in the obvious way.

Proof:
This follows immediately from

$$\{x \in X : (\sup_j f_j)(x) > a\} = \bigcup_j \{x \in X : f_j(x) > a\}$$

which is easily checked. (Do it.)
QED

Exercise: Would the above equality of sets hold if the $>$ in the two places were both replaced by $\geq$?

Proposition 4.8. *If $f_1, f_2, \dots$ is a sequence of measurable functions, then $\limsup_j f_j$ is a measurable function where this is of course defined by*

$$(\limsup_j f_j)(x) := \limsup_j (f_j(x)).$$

In particular, if (f_k) converges to the function f_∞ pointwise, then f_∞ is measurable.

Proof:
One notes first that

$$\limsup_j f_j = \inf_k (\sup_{n\geq k} f_n).$$

Applying the previous proposition twice, we obtain the result. The last statement follows as a special case.
QED

The following definition starts us off on our way.

Definition 4.9. *A* **simple** *function on* $(X, \mathcal{M})$ *is a function of the form*

$$f(x) = \sum_{i=1}^{n} c_i I_{E_i}$$

where $c_1, \ldots, c_n$ *are real numbers,* $E_1, \ldots, E_n$ *are disjoint sets in* $\mathcal{M}$ *and* I_{E_i} *is the indicator function of* E_i *which means it is 1 on* E_i *and 0 otherwise.*

Note that the indicator function of the rationals, which was not Riemann integrable, is simple in this sense, even with n being 1 (despite the fact that it looks sort of complicated!) Simple functions can approximate any measurable function in a pointwise sense as the following theorem indicates.

Theorem 4.10. *(Folland Theorem 2.10) If* $(X, \mathcal{M})$ *is a measurable space and* $f : X \to [0, \infty]$ *is measurable, then there exists a sequence* (ϕ_n) *of simple functions such that* $0 \leq \phi_1 \leq \phi_2 \leq \ldots$ *so that* ϕ_n *approaches* f *pointwise. Moreoever* ϕ_n *approaches* f *uniformly on any set where* f *is bounded.*

Proof:
We don't do the proof. It is a good exercise to think how you would do it; the idea is to break up the y-axis rather than the x-axis. The details are in F if you don't manage to do it on your own.
QED

4.3 The Lebesgue Integral and the main convergence theorems: Monotone Convergence Theorem, Fatou's Lemma and the Lebesgue Dominated Convergence Theorem

We now finally begin to introduce the Lebesgue integral which we will do in steps. Our measure space is $(X, \mathcal{M}, m)$.

We first need to consider the class of nonnegative measurable functions

$$L^+((X,\mathcal{M},m)) := \{f : X \to [0,\infty], f \text{ is measurable}\}$$

Definition of the integral for nonnegative simple functions

Definition 4.11. *If ϕ is a simple function in $L^+((X,\mathcal{M},m))$,*

$$\phi(x) = \sum_{i=1}^{n} c_i I_{E_i} \;\; (c_i \geq 0 \, \forall i),$$

then we define the integral of ϕ by

$$\int \phi(x)\; dm(x) := \sum_{i=1}^{n} c_i m(E_i).$$

It is "obvious" this is the right definition. But one needs to check that the integral is independent of the representation since a simple function can be written as a sum in more than one way. (Playing with a few examples will convince you that this is fine).

Definition 4.12. *If ϕ is a simple function in $L^+((X,\mathcal{M},m))$ and $A \in \mathcal{M}$,*

$$\text{we define } \int_A \phi(x)\; dm(x) := \int \phi(x) I_A \; dm(x)$$

(noting that the RHS is a simple function and hence the integral of the RHS is defined).

The following gives some simple properties of our integral for simple functions. The first three parts should be viewed as "obvious".

Proposition 4.13. *(Proposition 2.13 in Folland)*
Let ϕ and ψ be simple nonnegative functions. Then the following hold.
a. $\int c\phi(x)\; dm(x) = c \int \phi(x)\; dm(x) \;\; \forall c \geq 0$.
b. $\int (\phi(x) + \psi(x))\; dm(x) = \int \phi(x)\; dm(x) + \int \psi(x)\; dm(x)$.
c. $\phi \leq \psi$ *implies that* $\int \phi(x)\; dm(x) \leq \int \psi(x)\; dm(x)$.
d. The mapping from $\mathcal{M}$ to $[0,\infty]$ given by $A \to \int_A \phi(x)\; dm(x)$ is a measure on $\mathcal{M}$.

(Note a and b say that integration is "linear" for nonnegative simple functions. d will return to us when we prove the Radon-Nikodym Theorem much later on.)

Proof:
a-c are pretty straightforward and so we just do d. Fix a simple function $\phi(x) = \sum_{i=1}^n c_i I_{E_i}$ with E_i's disjoint and $c_i \geq 0$ for all i. Let $A_1, A_2, \ldots$ be disjoint elements in $\mathcal{M}$ and let $A := \bigcup_i A_i$. Observe that for any $B \in \mathcal{M}$, one has

$$\phi I_B = \sum_{i=1}^n c_i I_{E_i \cap B} \text{ and hence}$$

$$\int_A \phi \, dm = \int \phi I_A \, dm = \int \sum_{i=1}^n c_i I_{E_i \cap A} dm = \sum_{i=1}^n c_i m(E_i \cap A) = \sum_{i=1}^n c_i \sum_{j=1}^\infty m(E_i \cap A_j) =$$

$$\sum_{j=1}^\infty \sum_{i=1}^n c_i m(E_i \cap A_j) = \sum_{j=1}^\infty \int \sum_{i=1}^n c_i I_{E_i \cap A_j} \, dm = \sum_{j=1}^\infty \int \phi I_{A_j} \, dm = \sum_{j=1}^\infty \int_{A_j} \phi \, dm.$$

QED

We now define the integral for all functions in $L^+((X, \mathcal{M}, m))$.

Definition 4.14. *(Definition of the integral for nonnegative measurable functions)*
If $f \in L^+((X, \mathcal{M}, m))$, we define

$$\int f(x) dm(x) := \sup\{\int \phi \, dm : 0 \leq \phi \leq f, \ \phi \text{ simple } \}.$$

Note this integral is certainly allowed to be ∞. It is easy to see (check this) that if $f, g \in L^+((X, \mathcal{M}, m))$, then

$$f \leq g \rightarrow \int f dm \leq \int g dm \tag{9}$$

and for $c \geq 0$

$$\int cf dm = c \int f dm.$$

We would think that integration should be a *linear* operation and in fact, it is true that with f, g as above, one has

$$\int (f+g) dm = \int f dm + \int g dm. \tag{10}$$

While this is true, we will need to develop our first convergence theorem to verify this; however the $\geq$ in (10) can be established fairly easily.

Exercise: Prove $\geq$ in (10).

We now come to our first convergence theorem. It is the key step leading to all of our convergence theorems.

Theorem 4.15. *(Monotone Convergence Theorem) Given a measure space $(X, \mathcal{M}, m)$, let (f_n) be a sequence of functions in $L^+((X, \mathcal{M}, m))$ satisfying*

$$0 \leq f_1 \leq f_2 \leq f_3 \ldots (\textit{meaning that these inequalities hold for every } x)$$

and define f by

$$f(x) := \lim_{n\to\infty} f_n(x) = \sup_n f_n(x).$$

Then

$$\int f dm = \lim_{n\to\infty} \int f_n dm. \textit{ The two sides might certainly be } \infty$$

Proof:
By (9), $\int f_n dm$ is increasing and hence has a limit. For the same reason $\int f dm \geq \int f_n dm$ for every n and hence

$$\int f dm \geq \lim_{n\to\infty} \int f_n dm.$$

(Now the real argument starts; it will use measure theory and the definition of the integral.) To prove the reverse inequality, we need to show, according to the definition of $\int f dm$, that for every simple function $\phi \leq f$,

$$\int \phi \, dm \leq \lim_{n\to\infty} \int f_n dm. \tag{11}$$

Fix such a ϕ. Now let $\alpha < 1$ and define

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

On observes that $E_1 \subseteq E_2 \subseteq E_3 \ldots$ and $X = \bigcup_n E_n$. (To see the latter claim, it is simplest to consider two cases: $f(x) = 0$ and $f(x) > 0$.) We now have for every n

$$\int f_n dm \geq \int f_n I_{E_n} dm \geq \int \alpha\phi I_{E_n} dm = \alpha \int_{E_n} \phi \, dm$$

Let $n \to \infty$ and use Proposition 4.13(d) together with continuity from below for measures to obtain

$$\lim_{n\to\infty} \int f_n dm \geq \alpha \int \phi\, dm.$$

Since this inequality holds for every $\alpha < 1$, we obtain (11).
QED

Before we continue to our two other major converence theorem results (Fatou's Lemma and the Lebesgue Dominated Convergence Theorem), we give two corollaries of the Monotone Convergence Theorem.

Corollary 4.16. *(Linearity) If f_1 and f_2 and in $L^+((X, \mathcal{M}, m))$, then*

$$\int (f+g) dm = \int f dm + \int g dm.$$

Proof:
By Theorem 4.10, choose ϕ_n and ψ_n to be simple functions increasing upward to f_1 and f_2 respectively. Then $\phi_n + \psi_n$ is a sequence of simple functions increasing upward to $f_1 + f_2$. This gives

$$\int f_1 + f_2\, dm = \lim_{n\to\infty} \int \phi_n + \psi_n\, dm = \lim_{n\to\infty} \int \phi_n + \int \psi_n dm = \int f_1 + \int f_2$$

where the MCT was used in the outer most equalities and Proposition 4.13(b) was used in the middle equality.
QED

Corollary 4.17. *If $f_1, f_2 \ldots$ in $L^+((X, \mathcal{M}, m))$, then*

$$\int (\sum_{i=1}^{\infty} f_i) dm = \sum_{i=1}^{\infty} (\int f_i dm)$$

Proof:
By the previous corollary, we have that for every N,

$$\int (\sum_{i=1}^{N} f_i) dm = \sum_{i=1}^{N} (\int f_i dm).$$

We want to let $N \to \infty$. The RHS goes to the RHS further up (just by definition of an infinite sum) while the LHS goes to the LHS further up by

the MCT.
QED

We prove one more elementary fact before moving on to our other convergence theorems.

Proposition 4.18. *If $f \in L^+((X, \mathcal{M}, m))$, then*

$$\int f dm = 0 \text{ if and only if } f = 0 \text{ a.e.}$$

Proof:
The "if" direction is easier and left to the reader. For the "only if" direction, if we assume the RHS does not hold, then we have

$$m(x : f(x) > 0) > 0.$$

Let $E_n := \{x : f(x) > 1/n\}$ and observe that $E_1 \subseteq E_2 \subseteq E_3 \ldots$ and $\bigcup_n E_n = \{x : f(x) > 0\}$. Since $m(\bigcup_n E_n) > 0$, continuity from below yields that there exists N with $m(E_N) > 0$.

Now consider the nonnegative simple function

$$\phi := \frac{1}{N} I_{E_N}.$$

We have $\phi \leq f$ and so

$$\int f \, dm \geq \int \phi dm = \frac{1}{N} m(E_N) > 0.$$

QED

An alternative and perhaps easier proof of the "only if" direction is to use Markov's inequality which comes later in this chapter.

We now move to our second convergence theorem.

Theorem 4.19. *(Fatou's Lemma)*
If (f_n) is a sequence of functions in $L^+((X, \mathcal{M}, m))$, then

$$\int \liminf_{n \to \infty} f_n \, dm \leq \liminf_{n \to \infty} \int f_n \, dm.$$

Proof:
Fix an integer k. Now for all $j \geq k$, we have

$$\inf_{n \geq k} f_n \leq f_j$$

and hence

$$\int \inf_{n \geq k} f_n \leq \int f_j.$$

Since this is true for all $j \geq k$, we have

$$\int \inf_{n \geq k} f_n dm \leq \liminf_{n \to \infty} \int f_n dm. \tag{12}$$

We have what we want on the RHS and now we take $k \to \infty$. Note that $\inf_{n \geq k} f_n$ is an increasing sequence in k and converges to $\liminf f_n$. Hence by the MCT, the LHS in (12) converges, as $k \to \infty$, to $\int \liminf_{n \to \infty} f_n dm$, completing the proof.
QED

Remark:
1. Of course if $f_n \in L^+((X, \mathcal{M}, m))$ converges to some function f pointwise, this simply yields

$$\int f dm \leq \liminf_{n \to \infty} \int f_n dm.$$

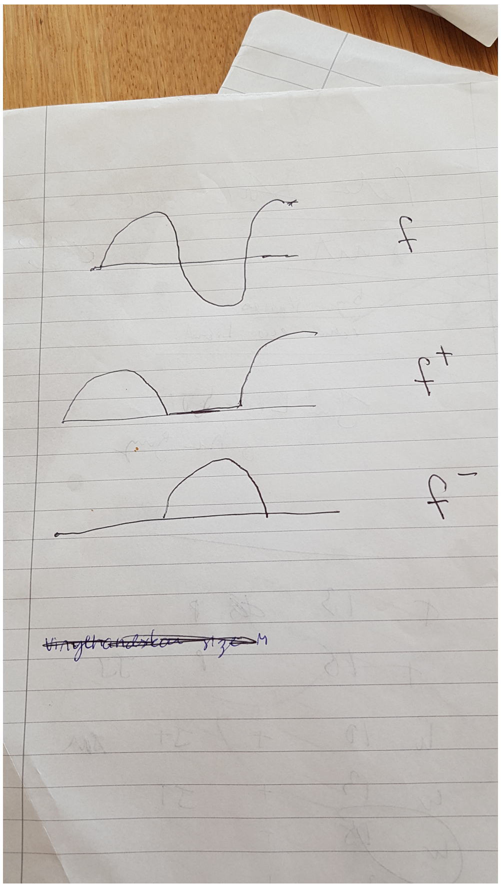
2. We do not necessarily have equality in Fatou's Lemma even in the special case of the previous line. On $[0, 1]$ with Lebesgue measure, letting $f_n = nI_{(0,1/n]}$, one notes that f_n converges to the function 0 pointwise, but $\int f_n dm = 1$ for every n.
3. Fatou's Lemma is very often used in analysis.
4. For finite measure spaces, uniform convergence would suffice for convergence of the integrals but not for infinite measure spaces. However, we will prove something much stronger than this.

Before moving on to the important Lebesgue Dominated Convergence Theorem, we need to define the integral for general functions, not only for nonnegative functions, which we do now.

We need to break up a general function into its so-called *positive and negative parts* which are defined as follows (see picture).

$$f^+(x) = \max\{f(x), 0\}, \ \ f^-(x) = \max\{-f(x), 0\}.$$

Note that both f^+ and f^- are nonnegative, $f = f^+ - f^-$ and $|f| = f^+ + f^-$.



Definition 4.20. *(Definition of the integral for* general *measurable functions) If* $f : (X, \mathcal{M}, m) \to \overline{R}$*, define*

$$\int f(x)dm(x) := \int f^+(x)dm(x) - \int f^-(x)dm(x)$$

provided that at least one of the two terms on the RHS is finite. (Otherwise, the integral is not defined).

Note that the integral can be ∞ or $-\infty$ but we avoid $\infty - \infty$ by requiring at least one of the two integrals to be finite.

Definition 4.21. *If* f *is a measurable function on* $(X, \mathcal{M}, m)$ *and* $\int f dm$ *is defined and* finite*, we say that* f *is* **Lebesgue integrable***. (This is the same as having* $\int |f| dm$ *being finite.)*

Notation: We let

$$L^1((X, \mathcal{M}, m)) := \{f : (X, \mathcal{M}, m) \to \overline{R} : \int |f| dm < \infty\}$$

and more generally, for $p \geq 1$, we let

$$L^p((X,\mathcal{M},m)) := \{f : (X,\mathcal{M},m) \to \overline{R} : \int |f|^p dm < \infty\}.$$

These so called " L^p spaces" can be made into so-called Banach spaces and L^2 into a Hilbert space but we certainly will not get into this here.

Example: Is $f(x) = \sin x/x$ integrable on $(0,\infty)$ with Lebesgue measure?
It turns out that

$$\lim_{N\to\infty}\int_0^N \frac{\sin x}{x}dx \text{ exists and is finite (and even is } \pi/2).$$

Does that answer our question?
No. $\sin x/x$ is not integrable on $(0,\infty)$ since one can check that

$$\int_0^\infty |\frac{\sin x}{x}|dx = \infty.$$

The integrals of both the positive and negative parts are infinite but some cancellation allows the earlier integral to converge. It is analogous to the alternating series

$$\sum_{n=1}^\infty \frac{(-1)^n}{n}$$

which converges but not absolutely. (So called conditional but not absolute convergence requires a "predecided order" on the domain ($[0,\infty)$ in the previous case, N in this case).)

Remark: As one would expect, integration is linear for functions in L^1. This is proved in Proposition 2.21 in F. It it not so interesting and basically, one just uses the definition of the general integral, linearity of the integral for positive functions and moves things around. We will use this linearity in our next theorem.

Theorem 4.22. *(Lebesgue Dominated Convergence Theorem) Let (f_n) be a sequence of functions in $L^1((X,\mathcal{M},m))$ which converges pointwise to a function f. Assume that there exists $g \in L^1((X,\mathcal{M},m))$ such that for all n*

$$|f_n| \leq g.$$

Then $f \in L^1((X,\mathcal{M},m))$ and

$$\int f dm = \lim_{n\to\infty}\int f_n dm.$$

Proof:
Since $|f_n| \leq g$ for all n and $f_n \to f$, we also have $|f| \leq g$ and hence $f \in L^1((X, \mathcal{M}, m))$. Observe that for all n

$$g + f_n \geq 0 \text{ and } g - f_n \geq 0 \text{ and hence}$$

$$g + f \geq 0 \text{ and } g - f \geq 0.$$

Applying Fatou's Lemma to $(g + f_n)$ (and using linearity twice), we get

$$\int g\, dm + \int f dm = \int g + f\, dm \leq \liminf_{n\to\infty} \int g + f_n dm = \int g\, dm + \liminf_{n\to\infty} \int f_n dm.$$

Subtracting $\int gdm$ from both sides gives

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

Similarly, we have

$$\int g\, dm - \int f dm = \int g - f dm \leq \liminf_{n\to\infty} \int g - f_n\, dm = \int g\, dm - \limsup_{n\to\infty} \int f_n\, dm$$

(Why did the lim inf becomes a lim sup?) Subtracting $\int gdm$ from both sides gives

$$\int f dm \geq \limsup \int f_n dm.$$

So we have

$$\int f dm \leq \liminf \int f_n dm \leq \limsup \int f_n dm \leq \int f dm.$$

Hence the limit exists and is $\int f dm$ as claimed.
QED

Exercises:

1. Find an example of functions (f_n) with $0 \leq f_n \leq 1$ for all n, f_n converges to 0 uniformly but $\int f_n$ does not go to 0. Why can you not apply the LDC Theorem here? (Hint: You will need to use an infinite measure space.)

2. Find an example of nonnegative functions (f_n) with f_n converging to 0 uniformly, $\int f_n$ goes to 0 but such that there does not exist a function g as in the assumptions of the LDC Theorem. (Hint: You will need to use an infinite measure space.)

3. Find an example of functions (f_n) on a finite measure space with f_n converging to 0 pointwise so that $\int f_n$ goes to 0 but such that there does not exist a function g as in the assumptions of the LDC Theorem.

The LDC can sometimes be used to justify the interchanging of differentiation and integration under some assumptions. For example, in justifying

$$\frac{d}{dt}\int_0^1 f(x,t)dx = \int_0^1 \frac{df(x,t)}{dt}dx$$

under some assumptions. Read Theorem 2.2.7 in F.

4.4 Different notions of convergence

In this subsection, we will introduce two natural notions of convergence. Before doing that, we first give a lemma.

Lemma 4.23. *If the sequence (f_n) and f are measurable functions on $(X, \mathcal{M}, m)$, then*

$$\{x : f_n(x) \to f(x)\} \in \mathcal{M}.$$

Proof:
Untangling what the definition of a limit is (and thinking a bit), it is not hard to see that the set above is the same as

$$\bigcap_{m=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\{x : |f_n(x) - f(x)| < 1/m\}.$$

This belongs to $\mathcal{M}$ since the events on the RHS do and then we are applying countable set operations.
QED

Exercise: Modify the proof of the above result to prove that

$$\{x : f_n(x) \text{ converges to a finite value }\} \in \mathcal{M}.$$

The following are important notions of convergence.

Definition 4.24. *If the sequence (f_n) and f are measurable functions on $(X, \mathcal{M}, m)$, then we say*
(i) f_n **converges to** *f* **a.e.** *if*

$m(\{x : f_n(x) \not\to f(x)\}) = 0$ *(note the previous lemma showed this set is in $\mathcal{M}$)*

(ii) f_n **converges to** *f* **in measure** *if for every $\epsilon > 0$,*

$$\lim_{n\to\infty} m(\{x : |f_n(x) - f(x)| \geq \epsilon\}) = 0.$$

Remarks:
1. These notions are central concepts. In the various convergence theorems that we have seen, like the LDC, one does not require convergence for all x, but only convergence a.e. In fact, one only requires convergence in measure in the LDC!
2. These two notions in probability theory are called “almost sure” convergence and “convergence in probability”.

Proposition 4.25. *1. There is an example where convergence a.e. occurs but not convergence in measure.*
2. Convergence a.e. implies convergence in measure if the measure space is finite.
3. Convergence in measure, does not imply convergence a.e. even if the measure space is finite.
4. Convergence in measure implies that there exists a subsequence for which one has convergence a.e.

Proof:
1. On $[0,\infty)$ with Lebesgue measure, let $f_n = I_{[n,n+1]}$. Check f_n goes to 0 for every x but not in measure.
2. Fix $\epsilon > 0$. Let

$$E_N = \{x : |f_n(x) - f(x)| \geq \epsilon \text{ some } n \geq N\}.$$

Observe that $E_1 \supseteq E_2 \supseteq E_3 \ldots$ and that

$$\bigcap_k E_k \subseteq \{x : f_n(x) \not\to f(x)\}$$

and hence by assumption $m(\bigcap_k E_k) = 0$. By continuity from above (which requires that the measure space be finite!), we get

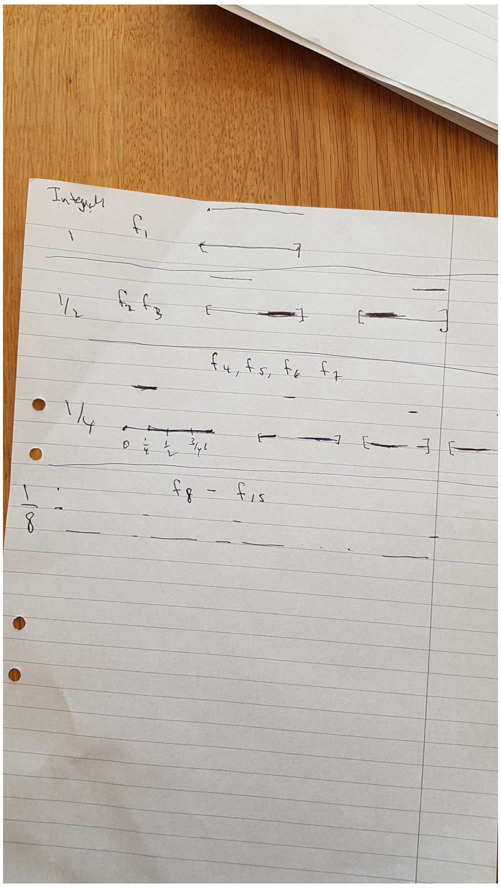
$$m(\{x : |f_N(x) - f(x)| \geq \epsilon\}) \leq m(E_N) \to 0 \text{ as } n \to \infty.$$

3. This is best described by a picture. See the (admittedly terrible) picture.
4. Assume (f_n) converges to f in measure. Then for each integer k, we can choose n_k so that

$$m(\{x : |f_{n_k}(x) - f(x)| \geq \frac{1}{k}\}) \leq \frac{1}{2^k}$$

and we can assume the n_k's are increasing in k. Letting

$$B_k := \{x : |f_{n_k}(x) - f(x)| \geq \frac{1}{k}\},$$



we have

$$\sum_k m(B_k) < \infty \text{ and hence from the Borel-Cantelli Lemma, we have}$$

$$m(B_k \text{ i.o.}) = 0.$$

Now, if x is not in (B_k i.o.), meaning $x \in B_k$ for only finitely many k, then $|f_{n_k}(x) - f(x)| \geq \frac{1}{k}$ for only finitely many k and hence

$$f_{n_k}(x) \to f(x).$$

QED

Exercise: Use the LDC Theorem and Proposition 4.25(d) to prove that the LDC Theorem is still true if we only assume convergence in measure.

The following theorem is often quoted. We might not have need for it and/or skip it but we state and prove it.

Theorem 4.26. *(Egoroff's Theorem)*
If (f_n) *is a sequence of measurable functions defined on a* finite *measure space* $(X, \mathcal{M}, m)$ *which converges to* f *pointwise (or a.e.). Then for all* $\epsilon > 0$*, there exists* $E \in \mathcal{M}$ *so that* $m(E) < \epsilon$ *and*

$$f_n \to f \text{ uniformly on } E^c.$$

Proof:
Let

$$E_n(k) := \{x : |f_m(x) - f(x)| \geq \frac{1}{k} \text{ for some } m \geq n\}$$

With k fixed, the above sets are decreasing in n and by assumption

$$\bigcap_n E_n(k) = \emptyset.$$

Since the measure space is finite, we have by continuity from above that

$$\lim_{n\to\infty} m(E_n(k)) = 0.$$

This is now been established for each k.

Now fix $\epsilon > 0$. For each k, choose n_k so that

$$m(E_{n_k}(k)) < \frac{\epsilon}{2^k}$$

and let

$$E := \bigcup_k E_{n_k}.$$

Note that $m(E) < \epsilon$ and that if $x \notin E$, then

$$|f_m(x) - f(x)| < \frac{1}{k} \text{ for all } m \geq n_k,$$

proving uniform convergence on E^c.
QED

4.5 Some inequalities: Markov and Chebyshev

Theorem 4.27. *(Markov's Inequality) Let* f *be a nonnegative measurable function on* $(X, \mathcal{M}, m)$*. Then for every* $\alpha > 0$*, one has*

$$m(\{x : f(x) \geq \alpha\}) \leq \frac{\int f dm}{\alpha}.$$

Proof:
We have

$$\int f dm = \int f I_{\{x:f(x)\geq\alpha\}} dm + \int f I_{\{x:f(x)<\alpha\}} dm \geq$$

$$\int \alpha I_{\{x:f(x)\geq\alpha\}} dm = \alpha m(\{x : f(x) \geq \alpha\}).$$

QED

Theorem 4.28. *(Chebyshev's Inequality) Let* f *be a measurable function on* $(X, \mathcal{M}, m)$ *with* $\int |f| dm < \infty$*. Then for any* $\alpha > 0$*, one has*

$$m(\{x : |f(x) - \int f dm| \geq \alpha\}) \leq \frac{\int (f - \int f\, dm)^2 dm}{\alpha^2}$$

Proof:
Apply Markov's inequality to the nonnegative function $(f(x) - \int f dm)^2$.
QED

Remark: Chebyshev's inequality is central and a key tool in probability theory.

5 Product Measures and The Fubini and Tonelli Theorems

5.1 Definition of a product measure

Goal: Given two measure spaces $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$, define a product of them. (The notion of a product exists for most mathematical objects, e.g. groups).

For the product space, we (obviously) should use as our set $X \times Y$. For the σ-algebra , we take it to be the smallest σ-algebra which contains the so-called "rectangles"

$$\mathcal{R} := \{A \times B : A \in \mathcal{M}, B \in \mathcal{N}\}$$

and we denote this σ-algebra $\sigma(\mathcal{R})$ by $\mathcal{M} \times \mathcal{N}$.

(Note for the Borel sets on $[0, 1]$, $\mathcal{R}$ contains many things which don't look like ordinary rectangles; e.g. $([0, \frac{1}{4}] \cup [\frac{5}{7}, \frac{6}{7}]) \times ([\frac{1}{3}, \frac{2}{3}] \cup \frac{10}{12}, \frac{11}{12}]) \in \mathcal{R}$).

It is quite clear that if we did have a "product" measure $\mu \times \nu$ on $\mathcal{M} \times \mathcal{N}$, it should satisfy $(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$ for all $A \in \mathcal{M}$ and $B \in \mathcal{N}$. The existence of a product measure is given by the following theorem.

Theorem 5.1. *(Existence of Product Measures) There exists a measure $\mu \times \nu$ on $(X \times Y, \mathcal{M} \times \mathcal{N})$ so that for all $A \times B \in \mathcal{R}$,*

$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B).$$

Moreover, $\mu \times \nu$ is the unique measure satisfying these properties if both μ and ν are σ-finite.

Proof outline:

Step 1. The collection of finite unions of disjoint rectangles is an algebra $\mathcal{A}$. To see this, one first shows/observes that the intersection of two elements in $\mathcal{R}$ is in $\mathcal{R}$ and that the complement of a set in $\mathcal{R}$ is the union of two elements in R. From here, one can either apply Proposition 1.7 in F or try to do the rest yourself.

Step 2. Define a function $(\mu \times \nu)_0$ on $\mathcal{A}$ by

$$(\mu \times \nu)_0(\bigcup_{i=1}^{n}(A_i \times B_i)) := \sum_{i=1}^{n} \mu(A_i)\nu(B_i).$$

One needs to check that this is well-defined since there is usually more than one way to represent a disjoint union of rectangles. The type of thing to check is that when one rectangle is a union of say four rectangles, then the above formula gives the same thing in both cases.

Step 3. $(\mu \times \nu)_0$ is a premeasure on $\mathcal{A}$.
The *key* step to prove this is to show that if we have a disjoint union

$$A \times B = \bigcup_i (A_i \times B_i) \quad \text{(an infinite union is allowed here)}$$

then

$$(\mu \times \nu)_0(A \times B) = \sum_i (\mu \times \nu)_0(A_i \times B_i). \tag{13}$$

It is then not so hard but tedious to check that (13) implies that $(\mu \times \nu)_0$ is a premeasure on $\mathcal{A}$. We now verify (13). For each $(x, y) \in X \times Y$, we have

$$I_A(x) \cdot I_B(y) = I_{A \times B}(x, y) = \sum_j I_{A_j \times B_j}(x, y) = \sum_j I_{A_j}(x) I_{B_j}(y).$$

Fix $y \in Y$ and integrate both sides with respect to x using μ. This yields, using MCT on the RHS

$$\mu(A) I_B(y) = \sum_j \mu(A_j) I_{B_j}(y).$$

Now integrate both sides with respect to y using ν, using MCT on the RHS to get

$$\mu(A)\nu(B) = \sum_j \mu(A_j)\nu(B_j).$$

Step 4. We can now apply Theorem 3.29 to complete the proof. Note that this also yields the uniqueness part.
QED

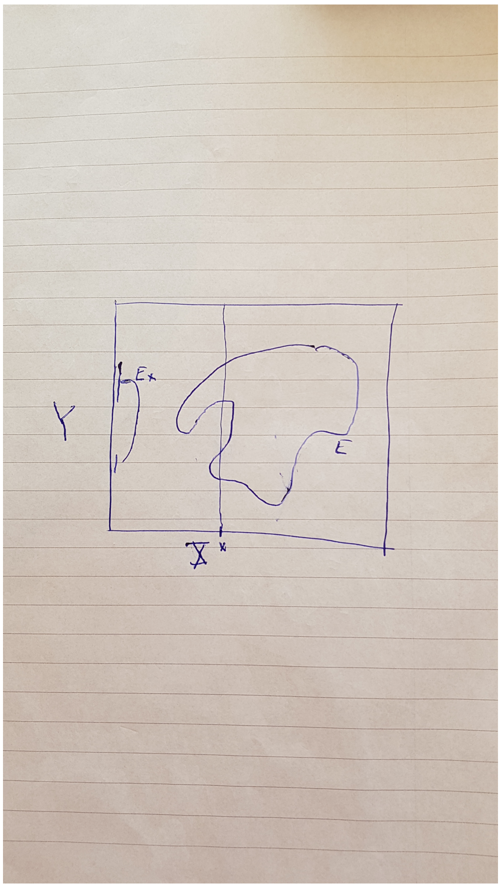
5.2 Sections/Projections

Definition 5.2. *If $E \subseteq X \times Y$, the x-***section** *of E (see picture) is*

$$E_x := \{y \in Y : (x, y) \in E\}$$

*and the y-***section** *of E is*

$$E^y := \{x \in X : (x, y) \in E\}.$$



Definition 5.3. *If $f : X \times Y \to R$, we let, for fixed $x \in X$,*

$$f_x : Y \to R \text{ given by } f_x(y) := f(x, y)$$

and for fixed $y \in Y$,

$$f^y : X \to R \text{ given by } f^y(x) := f(x, y).$$

Exercise: Check that if $f = I_E$, then $f_x = I_{E_x}$

Proposition 5.4. *(i). If $E \in \mathcal{M} \times \mathcal{N}$, then for all $x \in X$, $E_x \in \mathcal{N}$ and for all $y \in Y$, $E^y \in \mathcal{M}$.*
(ii). If $f : X \times Y \to R$ is $\mathcal{M} \times \mathcal{N}$-measurable, then for all $x \in X$, f_x is $\mathcal{N}$-measurable and for all $y \in Y$, f^y is $\mathcal{M}$-measurable.

Proof:
(i) The proof idea is to show that the sets E for which the statement is true

contains $\mathcal{R}$ and is a σ-algebra . Let

$$\mathcal{G} := \{E \subseteq X \times Y : \text{ the claim holds } \}$$

Now $\mathcal{R} \subseteq \mathcal{G}$ since

$$(A \times B)_x = \begin{cases} B & \text{if } x \in A \\ \emptyset & \text{if } x \notin A \end{cases}$$

and similarly for the other case. We next claim that $\mathcal{G}$ is a σ-algebra .

Exercise: Show $\mathcal{G}$ is a σ-algebra .
Hint: The two crucial but simple observations (which of course you check) are that

$$(E_x)^c = (E^c)_x \text{ and } \bigcup_i (E_i)_x = (\bigcup_i E_i)_x.$$

It now follows from these two claims that $\mathcal{G} \supseteq \sigma(\mathcal{R}) = \mathcal{M} \times \mathcal{N}$ as desired.

(ii) This follows immediately from (i) and the observation (of course which needs to be checked) that

$$(f_x)^{-1}(A) = (f^{-1}(A))_x.$$

QED

5.3 Fubini's Theorem for sets

Theorem 5.5. *(Fubini's Theorem for sets)*
Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces. If $E \in \mathcal{M} \times \mathcal{N}$, then the maps

$$x \to \nu(E_x) \text{ and } y \to \mu(E^y) \text{ (well defined by the previous result)}$$

are measurable functions on X and Y respectively. Furthermore

$$\mu \times \nu(E) = \int_X \nu(E_x) d\mu(x) = \int_Y \mu(E^y) d\nu(y).$$

Proof:
We first prove it where we assume the spaces are finite. By symmetry, it is enough to show that $x \to \nu(E_x)$ is measurable and $\mu \times \nu(E) = \int_X \nu(E_x) d\mu(x)$.

Let

$$\mathcal{C} := \{E \in \mathcal{M} \times \mathcal{N} : \text{ the above claim holds } \}.$$

The goal is to show that $\mathcal{C} = \mathcal{M} \times \mathcal{N}$.

STEP 1: $\mathcal{R} \subseteq \mathcal{C}$.

$$\nu((A \times B)_x) = \begin{cases} \nu(B) & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} = I_A(x)\nu(B)$$

Hence $x \to \nu((A \times B)_x)$ is measurable. In addition

$$\mu \times \nu(A \times B) = \int_X I_A(x)\nu(B)d\mu(x) = \int_X \nu((A \times B)_x)d\mu(x).$$

STEP 2: $\mathcal{C}$ is a $\mathcal{D}$-system.

We leave this to the reader. It is proved very similarly to Theorem 3.26. One needs to observe that $A \subseteq B$ implies that $(B \backslash A)_x = B_x \backslash A_x$ and to use the MCT.

Step 3. Using the fact that $\mathcal{R}$ is a π-system, Dynkin's $\pi - \lambda$ Theorem gives the second equality below and steps 1 and 2 give the containment.

$$\mathcal{M} \times \mathcal{N} = \sigma(\mathcal{R}) = \mathcal{D}(\mathcal{R}) \subseteq \mathcal{C}.$$

That proves the finite case.

Moving to the σ-finite case, let $X_1 \subseteq X_2 \subseteq \dots$ have finite measure with $X = \bigcup X_i$ and $Y_1 \subseteq Y_2 \subseteq \dots$ have finite measure with $Y = \bigcup Y_i$. Consider the measure $(\mu \times \nu)|_{X_n \times Y_n}$ which is a finite measure on $(X \times Y, \mathcal{M} \times \mathcal{N})$. By looking at rectangles, it is easy to see that

$$(\mu \times \nu)|_{X_n \times Y_n} = \mu|_{X_n} \times \nu|_{Y_n}.$$

By the finite case, for each $E \in \mathcal{M} \times \mathcal{N}$, the map $x \to \nu|_{Y_n}(E_x) = \nu(Y_n \cap E_x)$ is measurable on $(X, \mathcal{M})$ and

$$\mu \times \nu((X_n \times Y_n) \cap E) = (\mu \times \nu)|_{X_n \times Y_n}(E) = \int_X \nu|_{Y_n}(E_x)d\mu|_{X_n}(x) = \int_X I_{X_n}\nu(Y_n \cap E_x)d\mu(x).$$

Using continuity from below and letting $n \to \infty$, we conclude that $x \to \nu(E_x)$ is measurable on $(X, \mathcal{M})$. Also, letting $n \to \infty$ in the above display, using continuity from below and the MCT, we conclude that

$$(\mu \times \nu)(E) = \int_X \nu(E_x)d\mu(x),$$

as desired. We can of course now switch x and y.

QED

5.4 Counterexample without σ-finiteness

The following example shows that Theorem 5.5 without the σ-finite assumption is false. Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ each be $[0, 1]$ with the Borel sets with μ being Lebesgue measure and ν being counting measure (the latter not being σ-finite).

Let $D := \{(x, x) : x \in [0, 1]\}$ be the diagonal. It is easy to check that $D \in \mathcal{M} \times \mathcal{N}$ since D^c, which is open, can be expressed as a countable union of products of open intervals. (Why?)

Note $D_x = \{x\}$ for each $x \in X$ and $D^y = \{y\}$ for each $y \in Y$. Note that the map $x \to \nu(D_x)$ is the constant function 1 (which is measurable), and $y \to \mu(D^y)$ is the constant function 0 (which is measurable). However, we now have

$$\int_X \nu(D_x) d\mu(x) = 1 \neq 0 = \int_Y \mu(D^y) d\nu(y).$$

Moreover, we claim that

$$(\mu \times \nu)(D) = \infty$$

so that all three expressions in Theorem 5.5 are different.

First, what is the *definition* of $(\mu \times \nu)(D)$? Recalling its definition, we need to show that if the union of a countable set of rectangles $\{A_i \times B_i\}$ covers D, then $\sum_i \mu(A_i) \times \nu(B_i) = \infty$

To prove this, one needs to break up these rectangles into 3 types.

Type 1: Rectangles where B_i is finite
Type 2: Rectangles where B_i is infinite and A_i has 0 Lebesgue measure.
Type 3: Rectangles where B_i is infinite and A_i has positive Lebesgue measure.

One general observation to keep in mind is that for any collection of rectangles $S_i \times T_i$

$$\bigcup_i (S_i \times T_i) \subseteq (\bigcup_i S_i) \times (\bigcup_i T_i)$$

(with strict containment typically occurring).

Case 1: There exists a rectangle $A_i \times B_i$ of type 3. Then $\mu(A_i) \times \nu(B_i) = \infty$.

Case 2: There is no rectangle of type 3. In this case, we will show that the rectangles cannot in fact cover D, which then would complete the proof.

Assume that they do cover. Let B be the union of all the B_i's coming from type 1 rectangles. Clearly B is countable and hence the Lebesgue measure of B^c is 1.

By the general observation above, the union of all the type 1 rectangles is contained in $[0,1] \times B$. Therefore for each $x \notin B$, (x,x) must be contained in a type 2 rectangle $A_i \times B_i$. Hence B^c must be contained in the union of the A_i's coming from the type 2 rectangles. However, all such A_i's have Lebesgue measure 0 and hence so does their union. We have that B^c is contained in a set of Lebesgue measure 0 but we have seen it has Lebesgue measure 1. Contradiction.
QED

Remark: The fact that E_x is measurable for every measurable E did not require σ-finiteness. However, it turns out that without σ-finiteness, not only can Fubini's Theorem fail as we just saw, but it can also happen that the mapping $x \to \nu(E_x)$, which is well-defined, is not measurable. (An example of this was pointed out to me by David Fremlin.)

5.5 The Fubini and Tonelli Theorems

We now state the Fubini and Tonelli Theorems; Theorem 5.5 was an important special case and we obtain the general version by taking limits and doing a little work.

Theorem 5.6. *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces.*
(a) (Tonelli)
If $f \in L^+(X \times Y)$, then

$$g(x) := \int_Y f_x(y) d\nu(y) \in L^+(X)$$

and

$$\int_{X\times Y} f(x,y) d(\mu \times \nu)(x,y) = \int_X g(x) d\mu(x)$$

i.e.,

$$\int_{X\times Y} f(x,y) d(\mu \times \nu)(x,y) = \int_X \int_Y f_x(y) d\nu(y) d\mu(x)$$

(b) (Fubini)
If $f \in L^1(X \times Y)$, then

$$f_x \in L^1(Y, \mathcal{N}, \nu) \text{ for } \mu\text{-a.e. } x \in X.$$

Furthermore,

$$g(x) := \int_Y f_x(y) d\nu(y) \text{ (which is defined } \mu\text{-a.e.) belongs to } L^1(\mu)$$

and

$$\int_{X\times Y} f(x,y)d(\mu\times\nu)(x,y) = \int_X g(x)d\mu(x).$$

i.e.,

$$\int_{X\times Y} f(x,y)d(\mu\times\nu)(x,y) = \int_X\int_Y f_x(y)d\nu(y)d\mu(x)$$

(Same if we switch y and x.)

Outline of Proof:

(a) Unravelling definitions, the case where f is simply an indicator function is precisely what is covered by Theorem 5.5. The statements are closed under positive linear combinations and hence we have the result for all simple functions f. For a general $f \in L^+$, choose simple functions (f_n) converging upwards to f. Applying the result to each f_n and taking limits, everything goes through for f by using, among other things, MCT and sometimes using it twice. (Check the details.)

(b) Part (a) shows that if $f \in L^+ \cap L^1$, then

$$g(x) = \int_Y f_x(y)d\nu(y) < \infty \text{ for } \mu\text{-a.e. } x \text{ and } g \in L^1(\mu) \tag{14}$$

We now want to move from this to $f \in L^1$ in which case we write $f = f^+ - f^-$ with $f^+, f^- \in L^+ \cap L^1$. We then have from (14) that

$$g^+(x) := \int_Y f^+(x,y)d\nu(y) \text{ and } g^-(x) := \int_Y f^-(x,y)d\nu(y)$$

are each finite for μ-a.e. x and hence $f_x \in L^1(Y,\mathcal{N},\nu)$ for μ-a.e. $x \in X$, which is the first claim. We also have from (14) that both $g^+(x)$ and $g^-(x)$ are in $L^1(\mu)$ and hence so is

$$g^+(x) - g^-(x) = \int_Y f^+(x,y) - f^-(x,y)d\nu(y) = \int_Y f(x,y)d\nu(y),$$

which was the second claim. Finally, using (a),

$$\int_{X\times Y} f(x,y)d(\mu\times\nu)(x,y) = \int_{X\times Y} f^+(x,y)d(\mu\times\nu)(x,y) - \int_{X\times Y} f^-(x,y)d(\mu\times\nu)(x,y)$$

$$= \int_X\int_Y f_x^+(y)d\nu(y)d\mu(x) - \int_X\int_Y f_x^-(y)d\nu(y)d\mu(x) = \int_X\int_Y f_x(y)d\nu(y)d\mu(x)$$

by linearity twice.

QED

6 Weak and Strong Laws of Large Numbers in Probability Theory

6.1 The standard concepts from probability (including independence) placed into our measure theoretic context

Kolmogorov in around the 1930's placed probability theory on a firm mathematical basis; the mathematics which was used was measure and integration theory.

Definition 6.1. *A probability space is a measure space* $(\Omega, \mathcal{M}, P)$ *with* $P(\Omega) = 1$.

The philosophy or interpretations are as follows:
Ω is the set of "outcomes" of some "random" experiment.
$\mathcal{M}$ is the set of "events" to which we will assign a "probability". For $A \in \mathcal{M}$, $P(A)$ is the "probability that the event A occurs"; i.e., the probability that our "randomly chosen" $\omega \in \Omega$ falls in A.

So, $(\Omega, \mathcal{M}, P)$ describes or governs a "random experiment" where P tells us the "likelihood" that ω (chosen "randomly") falls in different sets.

Definition 6.2. *Given a probability space* $(\Omega, \mathcal{M}, P)$, *a* **random variable** *is a measurable real-valued function* X *on* $(\Omega, \mathcal{M}, P)$.

So a random variable is not really random as it is just a function. However, if ω is "random", then $X(\omega)$ is "random". Hence we call it a random variable.

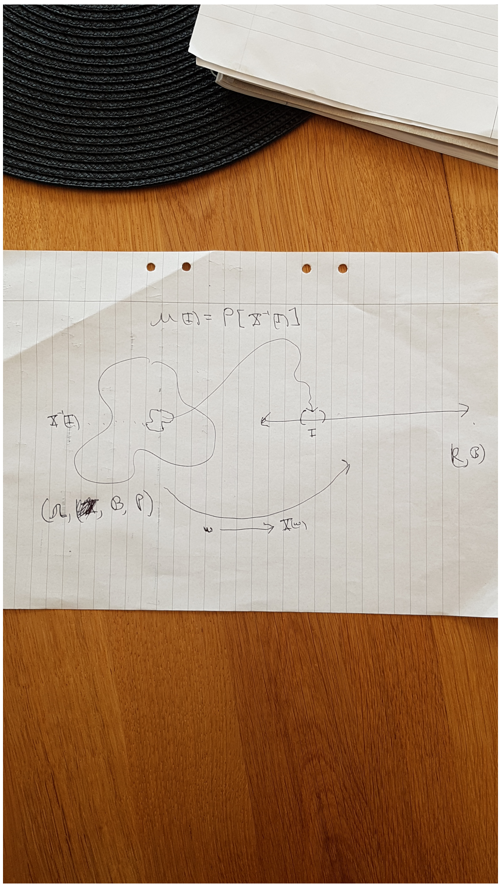
Definition 6.3. *If* X *is a random variable on a probability space* $(\Omega, \mathcal{M}, P)$, *its* **expectation**, *denoted* $E(X)$, *is simply defined by*

$$E(X) = \int X dP$$

provided this exists, meaning at least one of $\int X^+ dP$ *and* $\int X^- dP$ *is finite.*

Definition 6.4. *Given a random variable* X *on a probability space* $(\Omega, \mathcal{M}, P)$, *the* **distribution or law of X** *(see picture) is the probability measure* μ *on* $(R, \mathcal{B})$ *given by*

$$\mu(A) := P(X^{-1}(A)).$$



Exercise: Prove that μ above is a probability measure.

Remark: The distribution of X contains **all** the essential information of X.

Definition 6.5. *n random variables $X_1, X_2 \ldots, X_n$ on a probability space $(\Omega, \mathcal{M}, P)$ are called* **independent** *if for all Borel sets $B_1, B_2 \ldots, B_n$*

$$P(\bigcap_{i=1}^{n} X_i^{-1}(B_i)) = \prod_{i=1}^{n} P(X_i^{-1}(B_i)).$$

The notation is a little messy but the idea not hard: think carefully through it.

Definition 6.6. *An infinite collection of random variables on a probability space $(\Omega, \mathcal{M}, P)$ is called* **independent** *if each finite collection is independent as above.*

6.2 The Laws of Large Numbers

Laws of large numbers say things like "If we flip an infinite number of fair coins, we obtain half heads in the limit". However such a statement seems more philosophical than mathematical. Of course, we now turn this into mathematics.

Theorem 6.7. *(Weak and Strong Law of Large Numbers in a special case) Let $X_1, X_2, \ldots$ be independent random variables on some probability space such that the distribution of each X_i is $(\delta_1 + \delta_{-1})/2$; i.e., for all i,*

$$P(\{\omega : X_i(\omega) = 1\}) = P(\{\omega : X_i(\omega) = -1\}) = 1/2.$$

Then (WLLN)
(i).

$$\frac{S_n}{n} := \frac{\sum_{i=1}^n X_i}{n} \text{ converges in measure (in probability) } \to 0.$$

and (SLLN)
(ii).

$$\frac{S_n}{n} := \frac{\sum_{i=1}^n X_i}{n} \text{ converges almost everywhere (almost surely) } \to 0.$$

Remarks:
1. Since a.e. convergence implies convergence in measure on *finite* measure spaces, (ii) of course implies (i). Nonetheless, it is important to distinguish them (a) for historical reasons, (b) the proof of WLLN is a little simpler than that for SLLN and (3) the WLLN holds in certain cases where the SLLN does not hold. Also, both in fact hold under much weaker assumptions than those given here but we don't want to go into that in these notes.
2. The WLLN in this special case could be formulated in the 19th century since one just needed the concept of a limit of real numbers as each of the terms looked at concern a finite probability space. However, one did not have the conceptual framework in the 19th century to even state the SLLN.
3. It is good to ponder over which of these results is more natural from an applied or even philosophic point of view. I believe the WLLN is more natural than the SLLN for the following reason. The WLLN tells us that if you were to flip a coin a zillion times, then with very high probabilty, the percentage of heads will be in the interval $[.499, .5001]$. The SLLN tells us that if you were (for some reason!) to stand and flip a coin an infinite number of times, then with probability 1, the percentage of heads would converge to 1/2.

Proof:
We first compute $E(S_n^2)$. We have

$$E(S_n^2) = E((\sum_{i=1}^n X_i)(\sum_{j=1}^n X_j)) = \sum_{i,j=1}^n E(X_i X_j) = n + \sum_{i,j=1, i\neq j}^n E(X_i X_j).$$

Each X_i has expectation 0 (check!) and due to independence $E(X_i X_j) = 0$ when $i \neq j$ (convince yourself of this from first principles; you do not need to use a theorem from probability about the expectation of a product of independent random variables). We therefore obtain

$$E(S_n^2) = n$$

and hence, by linearity

$$E((S_n/n)^2) = 1/n.$$

Finally, fixing $\epsilon > 0$, we have

$$P(|S_n/n| \geq \epsilon) = P((S_n/n)^2 \geq \epsilon^2) \leq \frac{E((S_n/n)^2)}{\epsilon^2} = \frac{1}{n\epsilon^2}$$

where Markov's inequality was used for the only inequality above. Since the last term goes to 0 when $n \to \infty$ and ϵ is arbitrary, we have S_n/n goes to 0 in probability. This yields (i).

We now move to (ii) which is a little more involved. It turns out that it would be useful if we could show that for every $\epsilon > 0$

$$\sum_n P(|S_n/n| \geq \epsilon) < \infty. \tag{15}$$

Notice that the upper bound from (i) of $\frac{1}{n\epsilon^2}$ on the summands does not imply convergence since the harmonic series $\sum \frac{1}{n}$ diverges. Let's however not give up hope on proving (15) and let's for the moment see how we can prove (ii) **assuming** we can obtain (15) somehow.

Assuming (15), the Borel-Cantelli Lemma tells us that for every $\epsilon > 0$,

$$P(|S_n/n| \geq \epsilon \text{ i.o.}) = 0.$$

For integer k, letting $A_k := (|S_n/n| \geq 1/k \text{ i.o.})$, we therefore have $P(A_k) = 0$ and hence $P(\bigcup_k A_k) = 0$ by countable additivity. This is the same as $P(\bigcap_k A_k^c) = 1$. Finally, one observes (check this) that $\bigcap_k A_k^c$ is exactly the event that S_n/n converges to 0; i.e., $\omega \in \bigcap_k A_k^c$ if and only $S_n(\omega)/n$ converges to 0. Hence $P(S_n(\omega)/n \to 0) = 1$.

We now need to obtain (15) which we will prove by improving our estimates. The trick here is to use 4th moments rather than second moments as we had done in (i). We will show that we can replace the $E(S_n^2) = n$ above by

$$E(S_n^4) \leq 3n^2 \tag{16}$$

which gives $E((S_n/n)^4) \leq 3/n^2$. Assuming this, we fix $\epsilon > 0$ and we obtain similar to (i),

$$P(|S_n/n| \geq \epsilon) = P((S_n/n)^4 \geq \epsilon^4) \leq \frac{E((S_n/n)^4)}{\epsilon^4} \leq \frac{3}{n^2\epsilon^4}.$$

This yields (15) as desired.

So we are left with proving (16). Similar to before, we obtain

$$E(S_n^4) = \sum_{i,j,k,\ell} E(X_i X_j X_k X_\ell).$$

We break the index set into three groups, (a) $i = j = k = \ell$, (b) two of i, j, k, ℓ take one value and two take another value and (c) all other possibilities. Each of the terms in (a) or (b), which are of the form $E(X_i^4)$ and $E(X_i^2 X_j^2)$ respectively, are 1. Note that the terms in (c) include the case where i, j, k, ℓ are all distinct and things like $E(X_i^2 X_j X_k)$ or $E(X_i^3 X_j)$ where i, j, k are distinct and it is easy to check all of these terms are 0. (Check this.) Hence $E(S_n^4)$ is $n + n(n-1)3 \leq 3n^2$ once one combinatorially checks (it can be a little confusing) that the number of terms of type (b) is $n(n-1)3$.
QED

Remark: While we do not need it in these notes, we mention that, while we have seen much earlier that the converse of the Borel-Cantelli Lemma is not true, it **is** true that for *independent* events $\sum_i P(E_i) = \infty$ implies that $P(\limsup E_i) = 1$. This is even true under somewhat weaker conditions (for example pairwise independence) which is crucial for lots of applications.

We mentioned above that the strong law of large numbers holds under much weaker conditions. We will state but not prove the (almost) optimal result.

Theorem 6.8. *(Strong Law of Large Numbers: General case) Let $X_1, X_2, \ldots$ be independent random variables with the same distribution (on some probability space) with $E(|X|) < \infty$. Then*

$$\frac{S_n}{n} := \frac{\sum_{i=1}^n X_i}{n} \textit{ converges a.e. to } E(X).$$

6.3 Are our Laws of Large Numbers vacuous?

We have proved the laws of large numbers where there was an assumption that we had an infinite number of independent random variables. However, do we know that we can even have a probability space with an infinite number of independent variables? Maybe we cannot and our theorems are vacuous. Of course, fortunately one can (or otherwise I would be out of a job).

The point of this section is not to do things rigorously with all details but more to give a flavor of how this is done. I will mention two approaches.

Approach 1: Constructing an infinite product space.
We have seen how to construct the product of a finite number of measure spaces. It turns out that you can construct even an infinite product of probability spaces. (It is natural to stick with probability spaces since the total measure of a finite product is the product of the total measures of each component. In order to avoid getting 0 or ∞ in an infinite product, it is therefore natural to take our measures to have total measure 1.) The construction of such an object uses Theorem 3.29 where the algebra there will be sets describable by just finitely many coordinates. It takes some work to verify that the natural "measure" on this algebra is in fact a premeasure but can be carried out. Once one has an infinite product space, if we take a sequence of random variables where the nth one just depends on the nth coordinate of the infinite product space, then one can check that these random variables will be independent (this basically follows from the definition of a product space).

Approach 2: Using the unit interval with Lebesgue measure.
We already have a probability space all set up for us. Consider $([0,1], \mathcal{B}_{[0,1]}, m)$ where m is Lebesgue measure. Given $x \in [0,1]$, x has a binary expansion

$$x = \sum_{n=1}^{\infty} \frac{a_n(x)}{2^n}$$

where each $a_n(x) \in \{0,1\}$. This is unique except for countably many x's and so we can ignore this point. Now, for each $n \geq 1$, define the random variable

$$X_n(x) = 1 \text{ if } a_n(x) = 1 \text{ and } -1 \text{ if } a_n(x) = 0.$$

Exercise: Convince yourself that these are random variables and that these random variables $X_1, X_2, \ldots$ are independent and each has distribution $(\delta_1 + \delta_{-1})/2$.

Concerning the second approach, what does the SLLN say in this case? It says that a.e. point in $[0, 1]$ (with respect to Lebesgue measure) has asymptotically 1/2 1's in their binary expansion. This statement has nothing to do with probability and this statement would have made sense to Lebesgue even before probability theory was formalized.

6.4 Random Series: An interesting aside

The point of this subsection is to state an interesting special case of a theorem on random series.

Of course we all know that $\sum_n \frac{1}{n}$ diverges while $\sum_n \frac{(-1)^n}{n}$ converges. What happens if we put a random sign in front of $\frac{1}{n}$? In other words, let $\{X_n\}$ be independent random variables with $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$ for each n and we ask whether

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

converges. Well if all the X_n's are 1, then no and if they alternate, then yes. We should ask what happens a.e. (or a.s.) Interesting we have

Theorem 6.9. $\sum_{n=1}^{\infty} \frac{X_n}{n}$ *converges a.e.*

One can consider a slight variant by considering $\sum_n \frac{X_n}{n^\alpha}$ instead where $\alpha \in (0, 1)$ but the X_n's are the same as above. Again of course, if the X_n's are all 1's, then we have divergence and if the X_n's alternate, then we have convergence. But what do we have a.s.? One obtains the following interesting theorem showing that $\alpha = \frac{1}{2}$ is a "critical value" for the question.

Theorem 6.10. *(i) If* $\alpha > 1/2$*, then* $\sum_{n=1}^{\infty} \frac{X_n}{n^\alpha}$ *converges a.e.*
(ii) If $\alpha \in (0, 1/2]$*, then* $\sum_{n=1}^{\infty} \frac{X_n}{n^\alpha}$ *diverges a.e.. More specificially, one has that a.e.,* $\limsup_{n\to\infty} \sum_{k=1}^{n} \frac{X_k}{k^\alpha} = \infty$ *and* $\liminf_{n\to\infty} \sum_{k=1}^{n} \frac{X_k}{k^\alpha} = -\infty$.

What is happening that makes $\alpha = \frac{1}{2}$ so special? (Disclaimer: The following explanation uses words and concepts we have *not* done and so this is *not* self contained: the point is just to give a flavor of things).

Answer: It is immediate to check that the variance of $\sum_{k=1}^{n} \frac{X_k}{k^\alpha}$ is $\sum_{k=1}^{n} \frac{1}{k^{2\alpha}}$. Hence the sum of the variances of the summands is finite if $\alpha > 1/2$ and infinite if $\alpha \leq 1/2$. That seems like an important difference that could easily explain the above theorem. It is in fact the reason and certainly gives an explanation of why $\alpha = \frac{1}{2}$ is special. But it doesn't *prove* things.

For (i), one could start to argue as follows. The fact that the sum of the variances converges when $\alpha > 1/2$ implies that the sequence of partial sums $\sum_{k=1}^{n} \frac{X_k}{k^\alpha}$ (indexed by n) is a Cauchy sequence in L_2, the Hilbert space of square integrable functions. Since this is a complete space, this sequence must converge to something in L_2. L_2 convergence is stronger than convergence in measure (convergence in probability) and so one can conclude that these partial sums converge in probability to something. This still does not yield the a.e. convergence of the random sum. To obtain this, one needs to use an inequality by Kolmogorov, (not surprisingly called Kolmogorov inequality) which does the trick after some work.

For (ii), one needs to use a generalization of the Central Limit Theorem called the Lindenberg-Feller Central Limit Theorem. This theorem can be used to show that

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

converges *in distribution* to a standard normal random variable. Since the denominator approaches ∞, this precludes the a.e. convergence of the random series.

Part II

7 The Lebesgue-Radon-Nikodym Decomposition Theorem

In this second part, we will move into deeper aspects of measure theory. This chapter will provide a proof of the important Lebesgue-Radon-Nikodym Decomposition Theorem, which can be viewed as two theorems, namely the Radon-Nikodym Theorem and the Lebesgue Decomposition Theorem. Before that, we will first have to develop the so-called Hahn and Jordan Decompositions.

7.1 Signed measures, mutual singularity and the Hahn and Jordan Decomposition Theorems

Signed measures are something we introduce not so much for generalization (although there is motivation for that) but as a tool to analyze ordinary measures. First the definition.

Definition 7.1. *If* $(X, \mathcal{M})$ *is a measurable space, a* **signed measure** *is a map* ν *from* $\mathcal{M}$ *to* $[-\infty, \infty]$ *satisfying*
(i) $\nu(\emptyset) = 0$.
(ii) At most one of the values $\pm\infty$ *are assumed.*
(iii) If $A_1, A_2, \ldots,$ *are (pairwise) disjoint elements of* $\mathcal{M}$*, then*

$$\nu(\bigcup_i A_i) = \sum_i \nu(A_i)$$

where the sum on the RHS converges absolutely when the LHS is finite.

Remarks:
a. A measure is a signed measure.
b. If μ_1 and μ_2 are finite measures (or if at least one is a finite measure), then $\mu_1 - \mu_2$ is a signed measure. (Prove this!).
c. Condition (ii) is there to avoid having $\infty - \infty$. To see this more clearly, let $X := \mathbf{N}$ with all subsets being measurable. Each $k \in \mathbf{N}$ has (possibly negative) measure or weight $\nu(\{k\}) = a_k$. If $A \subseteq X$, we want to define

$$\nu(A) := \sum_{k \in A} a_k.$$

However, this will only make sense if either the sum of the positive terms or the sum of the negative terms is finite. (Conditional convergence makes no sense in this general context since that requires some arbitrary given ordering on the elements.)

One of the purposes of this subsection is to prove that every signed measure has the form $\mu_1 - \mu_2$ with μ_1 and μ_2 being measures with at least one being a finite measure. This is the so-called Jordan Decomposition Theorem which proves in fact something stronger, namely that μ_1 and μ_2 can also be taken to be **mutually singular**, a concept we will come to shortly. The hard work for this result is to first prove the so-called Hahn Decomposition Theorem which we will do first.

Definition 7.2. *If ν is a signed measure on $(X, \mathcal{M})$, a set $A \in \mathcal{M}$ is called* **a positive set** *if $\nu(B) \geq 0$ for all $B \subseteq A$ with $B \in \mathcal{M}$. A is called* **a negative set** *(***a null set***) if the $\geq$ is replaced by $\leq$ $(=)$. (Note that a set is null if and only if it is both a positive and a negative set.)*

Remark: A being a positive set is a strictly stronger statement than $\nu(A) \geq 0$. Exercise: Show this with an example.

Theorem 7.3. *(Hahn Decomposition Theorem) If ν is a signed measure on $(X, \mathcal{M})$, then X can be partitioned into two sets P, N ($P \cup N = X$, $P \cap N = \emptyset$) with $P, N \in \mathcal{M}$ where P is a positive set and N is a negative set. There is "almost uniqueness" in that if (P', N') is another such partition, then $P \triangle P'$ and $N \triangle N'$ are each null sets.*

Simple Example: Consider $([0,1], \mathcal{B}_{[0,1]})$ and we denote m by Lebesgue measure. Let

$$\nu(A) := m(A \cap [0, \frac{3}{4}]) - m(A \cap (\frac{3}{4}, 1]).$$

Then it is (hopefully) clear that a Hahn decomposition is given by $([0, \frac{3}{4}], (\frac{3}{4}, 1])$. In this case, the Hahn decomposition is essentially "given to you". The theorem says that such a decomposition always exists even though it might not be apparent from the start.

We separate the slightly harder part into a lemma. Before even this, we state two easy facts about signed measures which we won't prove; these are Propositions 3.1 and Lemma 3.2 in F.
(i) The continuity from below and from above for measures applies to signed measures as well.
(ii) A measurable subset of a positive set is a positive set (trivial) and a

countable union of (not necessarily disjoint) positive sets is a positive set (easy).

Lemma 7.4. *Let ν be a signed measure on $(X, \mathcal{M})$ which does not take the value ∞. If $\nu(A) > 0$, then there exists a measurable $B \subseteq A$ where $\nu(B) > 0$ and B is a positive set.*

Proof:
Observe that any set E with strictly positive ν-measure which is *not* a positive set contains a subset B with $\nu(B) > \nu(E)$. This is because if E is not a positive set, it contains a set C with strictly negative measure and then, by finite additivity, $E \backslash C$ would have measure strictly larger than that of E.

Now if the set A is a positive set, we are done. Otherwise, by the previous paragraph, we have subsets of A with strictly larger measure. We then let n_1 be the smallest possible integer so that there exists $A_1 \subseteq A$ so that $\nu(A_1) \geq \nu(A) + \frac{1}{n_1}$ and we then take any such set A_1.n If A_1 is a positive set, we are done. Otherwise, let n_2 be the smallest possible integer so that there exists $A_2 \subseteq A$ so that $\nu(A_2) \geq \nu(A_1) + \frac{1}{n_2}$ and we then take any such set A_2. If A_2 is a positive set, we are done. Otherwise, we continue inductively in the same way, constructing $(n_3, A_3), (n_4, A_4), \ldots$. The ν-measures of the A_n's are strictly increasing. If some A_k is a positive set, we stop then and we have what we want. Otherwise this continues forever and we define

$$A_\infty := \bigcap_k A_k.$$

We will eventually prove that A_∞ is a positive set. We first claim that the n_i's must go to ∞. To see this, note that for every k,

$$\nu(A_k) \geq \nu(A_{k-1}) + \frac{1}{n_k} \geq \ldots \geq \nu(A) + \sum_{i=1}^{k} \frac{1}{n_i}$$

and hence, by continuity from above and the fact that $\nu(A)$ is finite, we have

$$\nu(A_\infty) = \lim_{k\to\infty} \nu(A_k) \geq \nu(A) + \sum_{i=1}^{\infty} \frac{1}{n_i}.$$

Since $\nu(A_\infty) < \infty$, we obtain that the n_i's must go to ∞.

We now claim that A_∞ is a positive set. If not, there would again exist a subset B of A_∞ and an integer ℓ such that $\nu(B) \geq \nu(A_\infty) + \frac{1}{\ell}$. Since the

n_k's are going to ∞, we can choose $n_{k_0} > \ell$. This now contradicts the way A_{k_0} and n_{k_0} were chosen since we could have chosen B and ℓ instead of A_{k_0} and n_{k_0} since

$$\nu(B) \geq \nu(A_\infty) + \frac{1}{\ell} \geq \nu(A_{k_0-1}) + \frac{1}{\ell}.$$

QED

Proof of Theorem 7.3:
Assume WLOG ∞ is not obtained by ν. Let

$$m = \sup\{\nu(E) : E \text{ is a positive set}\}.$$

If $m = 0$, then Lemma 7.4 implies that X is a negative set (i.e., ν is a negative measure) and we would be done. Otherwise, we choose a sequence of positive sets (P_j) so that

$$\lim_{j\to\infty} \nu(P_j) = m.$$

Letting $P = \bigcup_j P_j$, we have that P is a positive set and we have $\nu(P) = m$. (Why?) (Note that this implies that $m < \infty$.) If we can show that P^c is a negative set, we would be done. However, if P^c is not a negative set, then there exists $E \subseteq P^c$ with $\nu(E) > 0$. By Lemma 7.4, E would then contain a measurable subset F which is a positive set and with $\nu(F) > 0$. This would imply that $P \cup F$ would be a positive set with ν-measure larger than m. This is a contradiction and hence we conclude that P^c is a negative set.

Finally, for the essential uniqueness, assume that (P', N') is another such partition. Since $P \backslash P' \subseteq P \cap N'$, $P \backslash P'$ is both a positive and a negative set and hence a null set. The same argument shows that $P' \backslash P$ is a null set and hence $P \triangle P'$ is a null set. For showing that $N \triangle N'$ is a null set one can either "apply the exact same argument" or better yet, observe that

$$N \triangle N' = P \triangle P'.$$

QED

The following definition is a very important definition. We state it for measures only. The definition for signed measures needs to be slightly modified.

Definition 7.5. *Two measures μ and ν on $(X, \mathcal{M})$ are* **mutually singular** *if X can be partitioned into two disjoint sets E and F in $\mathcal{M}$ so that $\mu(E) = 0 = \nu(F)$. (Loosely speaking, μ "lives on F" and ν "lives on E".) This is denoted by $\mu \perp \nu$.*

Exercises:
1. If μ and ν are measures on $(X, \mathcal{M})$ where X is a finite set and $\mathcal{M}$ consists of all subsets, characterize exactly when $\mu \perp \nu$.
2. Show that any two of the following measures on $[0, 1]$ are mutually singular. (i) An atomic measure with a countable number of atoms (ii) the Cantor measure and (iii) Lebesgue measure.
3. It is trivial to find an uncountable number of measures on $[0, 1]$ which are mutually singular, namely $\{\delta_x\}_{x \in [0,1]}$. However, one can in fact also find an uncountable number of measures on $[0, 1]$ which are mutually singular, none of which have an atom. Try to do this.
Hint: $[0, 1]$ can be represented by infinite binary sequences and one might consider the strong law of large numbers with the probability of a 1 varying, being any number $p \in [0, 1]$.

Theorem 7.6. *(Jordan Decomposition Theorem) If ν is a signed measure on $(X, \mathcal{M})$, then there exist unique measures ν^+ and ν^- so that ν^+ and ν^- are mutually singular and*

$$\nu = \nu^+ - \nu^-.$$

(The measure $\nu^+ + \nu^-$ is called the total variation of ν and is denoted by $|\nu|$; it is also important.)

Proof:
This is essentially a corollary of the Hahn Decomposition Theorem. Let P, N be a Hahn decomposition of ν. Let ν^+ be the "restriction of ν to P", meaning

$$\nu^+(A) := \nu(A \cap P)$$

and let ν^- be the "restriction of ν to N" but "reversed", meaning

$$\nu^-(A) := -\nu(A \cap N).$$

Then one easily checks that ν^+ and ν^- are mutually singular measures (do this!) with

$$\nu = \nu^+ - \nu^-.$$

We skip the uniqueness part; see Theorem 3.4 in F.
QED

The best way (in my opinion) to learn general concepts is to look at lots of examples.

Exercises:
1. Consider the signed measure ν on $([0,1], \mathcal{B}_{[0,1]})$ given by

$$\nu(A) = \int_A (2x-1)dx.$$

Find explicitly the Hahn and Jordan decompositions of ν.
2. Consider the signed measure ν on $([0,1], \mathcal{B}_{[0,1]})$ given by

$$\nu(A) = \int_A (\sin(2\pi x) - \cos(2\pi x))dx.$$

Find explicitly the Hahn and Jordan decompositions of ν.

7.2 Absolute continuity and the Lebesgue-Radon-Nikodym Decomposition Theorem

Besides mutual singularity of measures, the following concept will be another (perhaps even more) central concept in measure theory. We again give the definition only for measures but it extends fairly easily to signed measures; see F.

Definition 7.7. *Given two measures ν and μ on $(X, \mathcal{M})$, we say that ν is* **absolutely continuous** *with respect to μ, written $\nu \ll \mu$, if for all $A \in \mathcal{M}$,*

$$\mu(A) = 0 \textit{ implies that } \nu(A) = 0.$$

So it says that whenever μ gives a set measure 0, then ν also does. It does not mean that $\nu \leq \mu$. For example if μ is Lebesgue measure on $[0,1]$ and ν is two times Lebesgue measure on $[0,1]$, then we have $\nu \ll \mu$ as well as $\mu \ll \nu$.

The following is a simple but central example illustrating this concept. Consider a measure space $(X, \mathcal{M}, \mu)$ and a function $f \in L^+((X, \mathcal{M}, \mu))$. Define ν on $(X, \mathcal{M})$ by

$$\nu(A) := \int_A f(x)d\mu(x).$$

The fact that ν is a measure follows from linearity of the integral together with the MCT. Note that we already proved this in Proposition 4.13(d) in the special case that f is a simple function. Next, the fact that $\nu \ll \mu$ is immediate. We will abbreviate this measure by $f\mu$. Amazingly, any measure which is absolutely continuous with respect to μ has this form, subject to a σ-finiteness assumption. This is the content of the next theorem.

Theorem 7.8. *(The Radon-Nikodym Theorem) Let ν and μ be two measures on $(X, \mathcal{M})$ with $\nu \ll \mu$ and with ν and μ being σ-finite. Then there exists a measurable function $f_0 : (X, \mathcal{M}) \to [0, \infty)$ such that for all $A \in \mathcal{M}$,*

$$\nu(A) := \int_A f_0(x) d\mu(x).$$

Moreover, f_0 is unique in the sense that if g_0 is another such function, then

$$\mu\{x : f_0(x) \neq g_0(x)\} = 0.$$

(Of course, modifying f_0 on a set of μ-measure 0 still works.)

Remarks: 1. The f_0 above is called the Radon-Nikodym Derivative of ν with respect to μ.
2. If μ is Lebesgue measure on $(R, \mathcal{B})$ and ν is the distribution (or law) of a random variable which is absolutely continuous with respect to μ, then the Radon-Nikodym Derivative of ν with respect to μ is simply the "probability density function" from elementary probability.
3. The theorem is false in general without the σ-finite assumption on μ.
Exercise: Show this by considering the Borel sets on $[0, 1]$ with ν being Lebesgue measure and μ being counting measure.
4. It turns out however that σ-finiteness of ν is not actually needed as an assumption provided that f_0 is allowed to take the value ∞.

Proof:
We prove this under the assumption that both μ and ν are finite and just outline the extension to the σ-finite case, leaving the details of the extension to the reader. Define

$$\mathcal{F} := \{f : X \to [0, \infty) : \int_A f(x) d\mu(x) \leq \nu(A) \ \ \forall A \in \mathcal{M}\}.$$

Note $\mathcal{F}$ is nonempty since $f \equiv 0$ is in $\mathcal{F}$. Let

$$m := \sup\{\int f(x) d\mu(x) : f \in \mathcal{F}\}.$$

Note that $m \leq \nu(X)(< \infty)$.

claim: There exists $f_0 \in \mathcal{F}$ for which $\int f_0(x) d\mu(x) = m$; i.e. the supremum above is achieved.

A key observation in proving this is to show that if h_1, h_2 belong to $\mathcal{F}$, then $\max\{h_1, h_2\} \in \mathcal{F}$. One sees this by noting that for all $A \in \mathcal{M}$,

$$\int_A \max\{h_1, h_2\} d\mu(x) = \int_{A \cap \{h_1 \geq h_2\}} h_1(x) d\mu(x) + \int_{A \cap \{h_1 < h_2\}} h_2(x) d\mu(x)$$

$$\leq \nu(A \cap \{h_1 \geq h_2\}) + \nu(A \cap \{h_1 < h_2\}) = \nu(A).$$

Now choose $h_1, h_2, \ldots \subseteq \mathcal{F}$ so that

$$\lim_{n \to \infty} \int h_n(x) d\mu(x) = m$$

If we let

$$g_n := \max\{h_1, h_2, \ldots, h_n\}$$

we have that (1) each $g_n \in \mathcal{F}$ from the above, (2) $g_1 \leq g_2 \leq g_3 \ldots$ and

$$\lim_{n \to \infty} \int g_n(x) d\mu(x) = m$$

since $\int h_n(x) d\mu(x) \leq \int g_n(x) d\mu(x) \leq m$. Finally, letting

$$f_0 := \lim_{n \to \infty} g_n,$$

we have (why?) by MCT that $f_0 \in \mathcal{F}$ and that $\int f_0(x) d\mu(x) = m$, proving the **claim**.

Now, letting

$$\nu_0 := \nu - f_0 \mu$$

we have that ν_0 is a measure and our goal is to show that it is the zero measure. The idea is to show that if it were not the 0 measure, then we can push m up a bit, giving a contradiction.

Now, if it were the case that $\nu_0(X) > 0$, then there would exist $\epsilon > 0$ so that

$$\nu_0(X) - \epsilon \mu(X) > 0. \tag{17}$$

Let (P, N) be a Hahn decomposition for the signed measure $\nu_0 - \epsilon\mu$.

Case 1: $\mu(P) = 0$. Then, since $\nu \ll \mu$, we have that $\nu(P) = 0$ and hence $(\nu_0 - \epsilon\mu)(P) = 0$, contradicting (17).
Case 2: $\mu(P) > 0$. Note that

$$g_0 = f_0 + \epsilon I_P \in \mathcal{F} \quad \text{(this statement is also true in Case 1)}$$

since for all $A \in \mathcal{M}$

$$\int_A (f_0 + \epsilon I_P) d\mu(x) = \int_A f_0 d\mu(x) + \epsilon\mu(P \cap A) \leq \int_A f_0 d\mu(x) + \nu_0(P \cap A)$$

$$\leq \int_A f_0 d\mu(x) + \nu_0(A) = \nu(A).$$

Next, since $\mu(P) > 0$, we have that

$$\int g_0 d\mu(x) = m + \epsilon\mu(P) > m$$

contradicting the definition of m. Hence ν_0 is the zero measure and we are done.

For uniqueness, one notes that if $\mu\{x : f_0(x) \neq g_0(x)\} > 0$, then WLOG $\mu\{x : f_0(x) > g_0(x)\} > 0$ which yields

$$\int_{\{x: f_0(x) > g_0(x)\}} f_0(x) d\mu(x) > \int_{\{x: f_0(x) > g_0(x)\}} g_0(x) d\mu(x)$$

contradicting the fact that each integral equals $\nu\{x : f_0(x) > g_0(x)\}$.

Outline for the σ-finite extension:
For the σ-finite case, one can break up X into countably many pieces each of which has both finite μ and ν measure. Then one can do the above on each piece separately and put them altogether.
QED

Exercise: If μ and ν are measures on $(X, \mathcal{M})$ where X is a finite set and $\mathcal{M}$ consists of all subsets, characterize exactly when $\nu \ll \mu$ and describe the Radon-Nikodym derivative.

We will now move to the Lebesgue Decomposition Theorem. F combines the Radon-Nikodym Theorem and the Lebesgue Decomposition Theorem into one theorem. I find it clearer pedogogically to separate them although the proof of the second result will be heavily based on the proof of the first result. Again, we stick to measures instead of signed measures.

Theorem 7.9. *(Lebesgue Decomposition Theorem) Let ν and μ be two measures on $(X, \mathcal{M})$ with ν and μ being σ-finite. Then there exist unique measures ν_{ac} and ν_s so that*

$$\nu = \nu_{ac} + \nu_s$$

and

$$\nu_{ac} \ll \mu \text{ and } \nu_s \perp \mu.$$

In words, given two measures, one can break up the first into two pieces, one piece which is absolutely continuous with respect to the other measure and one piece which is singular with respect to the other measure.

Proof:
We do this only in the case that μ and ν is finite and we also leave the uniqueness to the reader; see F. We will modify the last part of the proof of the Radon-Nikodym Theorem and hence one must know the details of that.

We follow the proof of the RNT until and including the definition of ν_0. Now, rather than proving that ν_0 is the 0 measure, we will prove it is mutually singular with respect to μ; this would complete the proof with $\nu_{ac} := f_0\mu$ and $\nu_s := \nu_0$.

Let (ϵ_n) be a decreasing sequence of numbers in $(0, 1)$ converging to 0. Let (P_n, N_n) be a Hahn decomposition for the signed measure $\nu_0 - \epsilon_n\mu$. If there exists n with $\mu(P_n) > 0$, then we would be in Case (2) of the previous argument and that argument would lead to a contradiction. Hence we must have

$$\mu(P_n) = 0 \text{ for each } n.$$

Finally, let $P := \bigcup P_n$ and $N := \bigcap N_n$, which gives a partition of X. By countable additivity, $\mu(P) = 0$. Also, for each n, we have

$$\nu_0(N) \leq \nu_0(N_n) \leq \epsilon_n\mu(N_n) \leq \epsilon_n\mu(X)$$

from which it follows that $\nu_0(N) = 0$ by letting $n \to \infty$. Hence $\nu_0 \perp \mu$, as desired.
QED

There are a number of different proofs of the two previous theorems. We give here an alternative proof of LDT.

Alternative Proof of LDT using the statement of RNT but not its proof:
Clearly $\mu \ll \mu+\nu$ and we let f be the corresponding RN derivative. Partition X by defining $X_0 := \{x : f(x) = 0\}$ and $X_+ := \{x : f(x) > 0\}$ and then we decompose ν by

$$\nu = \nu|_{X_0} + \nu|_{X_+}$$

where these two measures are simply ν restricted to X_0 and X_+ respectively. Clearly $\nu|_{X_0}(X_+) = 0$ and $\mu(X_0) = 0$ since f vanishes on X_0. Hence $\nu|_{X_0} \perp \mu$.

We now need to show that $\nu|_{X_+} \ll \mu$. We have

$$\mu(A) = 0 \text{ implies } \int_A f d(\mu+\nu) = 0 \text{ implies } \int_A f d\nu = 0 \text{ implies } \int_{A\cap X_+} f d\nu = 0$$

$$\text{implies } \nu(A \cap X_+) = 0 \text{ implies } \nu|_{X_+}(A) = 0.$$

QED

Exercises:
1. If μ and ν are measures on $(X, \mathcal{M})$ where X is a finite set and $\mathcal{M}$ consists of all subsets, describe exactly the Lebesgue decomposition of ν with respect to μ and find the RN derivative for the absolutely continuous piece.
2. Consider the measurable space $(R, \mathcal{B})$ with the two finite measures μ and ν given below. (m denotes Lebesgue measure here).

$$\nu = \frac{2}{3}\delta_{-7} + 100\delta_4 + 3\delta_{5.5} + (x^2)m|_{[4,6]} \quad \mu = \frac{1}{4}\delta_4 + \frac{1}{2}\delta_{5.5} + \delta_{5.6} + (3x^3)m|_{[5,7]}.$$

Determine the Lebesgue decompositions of ν with respect to m (Lebesgue measure) and with respect to μ and find the RN derivatives for the absolutely continuous pieces in each case. (One needs to use the fact (see Proposition 3.9 in F) that if $m_1 \ll m_2$ with RN derivative f, then

$$\int g\, dm_1 = \int gf\, dm_2.$$

Note that when g is an indicator function, then this is just the definition of the RN derivative; one proves this by extending it to simply functions g by linearity and then all functions by taking limits.)

We mention here another important but fairly simple decomposition. Given any σ-finite measure space $(X, \mathcal{M}, \mu)$ with single points being measurable, we can always decompose μ into an atomic piece and a continuous piece as follows. If $\mathcal{A}$ is the set of atoms, we can write

$$\mu = \mu|_{\mathcal{A}} + \mu|_{\mathcal{A}^c}$$

and it is immediate to check that $\mu|_{\mathcal{A}}$ is atomic, $\mu|_{\mathcal{A}^c}$ is continuous and these measures are mutually singular.

We end this subsection with one other result which characterizes absolute continuity in many cases. It will also be used later on.

Proposition 7.10. *Let μ and ν be measures with ν finite. Then $\nu \ll \mu$ if and only if for every $\epsilon > 0$, there exists $\delta > 0$ so that*

$$\mu(A) < \delta \text{ implies that } \nu(A) < \epsilon.$$

Proof:
The "if" direction is essentially immediate (and does not require that ν be finite). Of course, check this.

The "only if" direction is more work and will use the fact that ν is finite. We prove the contrapositive. If the statement on the RHS fails, then there would exist an $\epsilon_0 > 0$ and sets (A_n) with $\mu(A_n) \le 1/2^n$ and $\nu(A_n) \ge \epsilon_0$. Now consider $A := \limsup A_n$. The Borel Cantelli Lemma tells us that $\mu(A) = 0$. On the other hand, $\nu(A) \ge \epsilon_0$ since for every n

$$\nu(\bigcup_{k=n}^{\infty} A_k) \ge \epsilon_0$$

and then we let $n \to \infty$ using continuity from above for ν which is valid since ν is a finite measure. This shows that it is not the case that $\nu \ll \mu$.
QED

Remark:
1. Consider the integers with all sets measurable. Let ν be counting measure and let $\mu(i) = 1/i$. Then $\nu \ll \mu$ but the condition on the RHS above fails. This shows ν being finite is a necessary condition.
2. The argument that $\nu(A) \ge \epsilon_0$ is basically a version of Fatou's Lemma, something which is called "Reverse Fatou" which states that for finite measure spaces

$$m(\limsup E_n) \ge \limsup_{n\to\infty} m(E_n).$$

7.3 What is happening on $(R, \mathcal{B})$?

It is insightful to see what happens in concrete cases. The following theorem describes what happens in the case of $(R, \mathcal{B})$.

Theorem 7.11. *Let μ be a measure on $(R, \mathcal{B})$ which is finite on compact sets. Then μ can be decomposed uniquely as*

$$\mu = \mu_d + \mu_{sc} + \mu_{ac}$$

where μ_d is an atomic measure ("d" for discrete), μ_{sc} is a continuous measure which is mutually singular with respect to Lebesgue measure and μ_{ac} is absolutely continuous with respect to Lebesgue measure.

Proof:
We saw near the end of the previous subsection that we can decompose μ into an atomic measure which we call μ_d and a continuous measure μ_c. We now apply the Lebesgue Decomposition Theorem with ν being μ_c and μ being Lebesgue measure in order to further decompose μ_c into $\mu_{sc}+\mu_{ac}$, completing

the proof.
QED

Remarks concerning **elementary probability**: In elementary probability courses, the distributions of random variables that occur usually either (1) only have an atomic part μ_d (so-called discrete random variables, like Poisson and geometric distributions) or (2) only have an absolutely continuous part μ_{ac} (so-called "continuous" random variables, like exponential and normal distributions). Sometimes, one might even have both of these parts. However, from the elementary probability level point of view, one is usually quite surprised that there are distributions that have a continuous singular part. Recall the Cantor measure is of this form which has only a continuous singular piece in the decomposition above. These would be random variables which are continuous, in the sense that there are no atoms, but which don't possess an ordinary probability density function (w.r.t. Lebesgue measure).

7.4 An outline of the proof of The Vitali–Hahn–Saks Theorem: An interesting Aside

We state the theorem.

Theorem 7.12. *Let (μ_n) be a sequence of probability measures on $(X, \mathcal{M})$ and assume that for each $E \in \mathcal{M}$, the limit*

$$\lim_{n\to\infty} \mu_n(E)$$

exists which we denote by $\mu_\infty(E)$. Then μ_∞ is a probability measure.

First, the Baire Category Theorem from Chapter 3 has many applications and one of them is the following which will be crucial for our proof of the above theorem.

Theorem 7.13. *Let (f_n) be a sequence of real-valued continuous functions on a complete metric space (Y, d) which converges to a function f pointwise. Then there exists a dense set $D \subseteq Y$ such that the family $\{f_n\}$ is equicontinuous at each point of D and as a consequence f is continuous at each point of D.*

Note that for clarify what the conclusion is saying is that for all $x \in D$ and $\epsilon > 0$, there exists a $\delta > 0$ so that for y with $d(x, y) < \delta$ we have that for all n

$$|f_n(x) - f_n(y)| < \epsilon.$$

In this subsection, we will give a sketch of the proof of Theorem 7.12. We will leave a number of details to the reader but all of the key ideas and steps will be discussed.

Step 1: One easily checks that given a sequence of probability measures ν_n, there exists a probability measure μ such that $\nu_n \ll \mu$ for all n.
Proof: Let $\mu := \sum_n \frac{1}{2^n}\nu_n$. (The fact that μ is a probability measure almost looks like the VHS Theorem but it is not and very easy to check.) QED

We need to introduce the crucial idea of a family of probability measures being uniformly absolutely continuous with respect to a probability measure.

Definition 7.14. *The family of probability measures $\{\nu_n\}$ is uniformly absolutely continuous with respect to the probability measure μ if for all $\epsilon > 0$, there exists $\delta > 0$ so that*

$$\mu(A) < \delta \textit{ implies that } \nu_n(A) < \epsilon \textit{ for all } n.$$

Step 2: A weak version of the VHS Theorem.

Theorem 7.15. *Let μ_n be a sequence of probability measures on $(X, \mathcal{M})$ and assume that for each $E \in \mathcal{M}$, the limit*

$$\lim_{n\to\infty} \mu_n(E) =: \mu_\infty(E)$$

exists. Assume that the family $\{\mu_n\}$ is uniformly absolutely continuous with respect to a probability measure μ. Then μ_∞ is a probability measure.

(The point is that this unif. abs. cont. is precisely what is needed to make the necessary change of limits.)

Proof:
Let $E_1, E_2, \ldots$ be disjoint and measurable. Since it is easy to check that μ_∞ is finitely additive, we have that for all n

$$\mu_\infty(\cup_i E_i) = \sum_{i=1}^{k} \mu_\infty(E_i) + \mu_\infty(\cup_{i\geq k+1} E_i).$$

The countable additivity is obtained if we show the last term is small for large k. Since μ is a probability measure it is certainly the case that $\mu(\cup_{i\geq k+1} E_i)$ is small for large k and hence by the u.a.c., $\mu_n(\cup_{i\geq k+1} E_i)$ is small for all n. Hence $\mu_\infty(\cup_{i\geq k+1} E_i)$ is small. QED

Step 3: If ν is absolutely continuous with respect to μ "at a fixed set" for a given ϵ, then it is absolutely continuous with respect to μ "at all sets" for that same ϵ.

Proposition 7.16. *Let μ and ν be probability measures on $(X, \mathcal{M})$ and assume that there is a measurable set E_0 and ϵ and δ so that for all measurable E*

$$\mu(E \triangle E_0) < \delta \text{ implies that } |\nu(E) - \nu(E_0)| < \epsilon.$$

Then for all measurable sets A and B, we have

$$\mu(A \triangle B) < \delta \text{ implies that } |\nu(A) - \nu(B)| < 2\epsilon.$$

Proof:
The key first step is to prove this when $B = \emptyset$. So we need to show that $\mu(A) < \delta$ implies that $\nu(A) < 2\epsilon$. Applying the first statement to $E = E_0 \backslash A$, we get $|\nu(E_0 \backslash A) - \nu(E_0)| < \epsilon$. Applying the first statement to $E = E_0 \cup A$, we get $|\nu(E_0 \cup A) - \nu(E_0)| < \epsilon$. It follows that

$$\nu(A) = \nu(A \cap E_0) + \nu(A \backslash E_0) < 2\epsilon.$$

Now for general A and B, by what we just proved, we have that $\mu(A \triangle B) < \delta$ implies that $\nu(A \triangle B) < 2\epsilon$. Next one easily observes that $\nu(A \triangle B) < 2\epsilon$ implies that $|\nu(A) - \nu(B)| < 2\epsilon$. QED

Step 4: Introduction of our complete metric space.
Define an equivalence relation on $\mathcal{M}$ by $A \sim B$ if $\mu(A \triangle B) = 0$. Define a metric d on $\mathcal{M}/\sim$ by

$$d(A, B) := \mu(A \triangle B).$$

Exercise: Show that (1) $\sim$ is an equivalence relation, (2) d is well defined, (3) d is a metric and (4) d is complete.
Hint: $(\mathcal{M}/\sim, d)$ embeds isometrically into $L^1(X, \mathcal{M}, \mu)$.

Step 5: Interpreting an absolutely continuous measure on $(X, \mathcal{M})$ as a function on $(\mathcal{M}/\sim, d)$.

Exercise: Show that if $\nu \ll \mu$ are probability measures, then we have a well defined map from $(\mathcal{M}/\sim, d)$ to R given by $A \to \nu(A)$. Show that this map is not well-defined if we do not have that $\nu \ll \mu$.

Proposition 7.17. *If $\nu \ll \mu$ are probability measures, then the above map is uniformly continuous.*

Proof: Letting $\epsilon > 0$, choose δ from Proposition 7.10. We can then apply Proposition 7.16 with $E_0 = \emptyset$ to obtain the desired conclusion. QED

Remark: In essence, Proposition 7.16 says that if one has continuity at some point, then one has uniform continuity.

Step 6: Proof of the VHS Theorem.

Given our sequence of probability measures μ_n, we can choose from Step 1 a probability measures μ so that $\mu_n \ll \mu$ for all n. Let f_n be the map associated to μ_n in Step 5. We have by the previous proposition that the f_n's are continuous on $(\mathcal{M}/\sim, d)$ and the main assumption is that these functions converge pointwise. Theorem 7.13 can now be applied to find some $E_0 \in \mathcal{M}/\sim$ so that the functions (f_n) are equicontinuous at E_0. Hence, given $\epsilon > 0$, there exists $\delta > 0$ so that the main assumption of Proposition 7.16 holds for E_0 and for each μ_n. Applying this proposition with $B = \emptyset$ allows us to conclude that the μ_n's are uniformly absolutely continuous with respect to μ and so we get the desired conclusion by Theorem 7.15. QED

8 Theory of Differentiation in R^n

We now will move into R^n rather than studying measure theory at a more abstract level. When one moves from the general to the more concrete, one can obtain deeper results but they require more work.

8.1 Historical comments and the first Lebesgue differentiation theorem

In the middle of the 19th century, mathematicians were attempting to prove that a continuous function must be differentiable at *some* point. Weierstrass then shocked the community when he constructed a continuous nowhere differentiable function on $[0, 1]$. It seems Bolzano also might have constructed such an object some years earlier. As a side comment, not only do continuous nowhere differentiable functions exist but most continuous functions are in fact nowhere differentiable in two different technical senses. First, most continuous functions are nowhere differentiable in a topological sense: the set of nowhere differentiable functions contains a dense G_δ (see Subsection 3.14). Secondly, most continuous functions are nowhere differentiable in a measure-theoretic or probabilistic sense: there is a natural probability measure on the space of continuous functions (which is called Brownian Motion and is the most important process in probability theory) and it is the case that almost every function (with respect to this measure) is nowhere differentiable.

Given these crazy functions, it is reported to have been a relief when Lebesgue proved that, on the other hand, increasing functions behave much more nicely.

Theorem 8.1. *(Lebesgue) Let $f : [0, 1] \to R$ be monotone ($x \leq y$ implies that $f(x) \leq f(y)$), then for a.e. x, f is differentiable with a finite derivative.*

We will follow F where a more general version concerning measures in R^n is obtained and then the above will be a corollary. This will involve, among other things, a so-called covering lemma and the so-called Hardy Littlewood maximal function.

8.2 The fundamental theorem of calculus

In a certain sense, most of the rest of notes concerns the question of how one can generalize the fundamental theorem of calculus. Let's first state the usual versions of this which are given in calculus classes.

Theorem 8.2. *1. (First fundamental theorem of calculus). If f is a continuous function on $[a, b]$ and*

$$F(x) := \int_a^x f(t)\ dt,$$

then $F'(x) = f(x)$ for all x.
2. (Second fundamental theorem of calculus). If f is a continuously differentiable function on $[a, b]$, then

$$\int_a^b f'(x)\ dx = f(b) - f(a).$$

We first discuss the first fundamental theorem. First, we will argue that this cannot possibly hold for all measurable bounded functions. Let f be a continuous function on $[a, b]$ and let g be any measurable function which agrees with f a.e. We then have that $\int_a^x f(t)\ dt = \int_a^x g(t)\ dt$ for every x. Hence we get for every x

$$(\int_a^x g(t)\ dt)'(x) = f(x)$$

and therefore

$$(\int_a^x g(t)\ dt)'(x) = g(x)$$

will fail at every point x where $f(x) \neq g(x)$. Nonetheless, observe that in this case we do have, since $f = g$ a.e.,

$$(\int_a^x g(t)\ dt)' = g(x) \text{ a.e.}$$

So the *most* one can hope for in general, say for bounded measurable f, is that the first fundamental theorem of calculus holds at a.e. x. We will in fact show this and even obtain a version of this in higher dimensions. This higher dimensional analogue is called Lebesgue's Differentiation Theorem, see Theorem 8.10. We mention that even in 1-d dimension not only can it happen that $(\int_a^x g(t)\ dt)'(x) \neq g(x)$ but it can also happen that the indefinite integral $\int_a^x g(t)\ dt$ is not differentiable at certain points. However, a.e., it will be differentiable AND the derivative will agree with g. This starts to sound a bit like Theorem 8.1. However the proof of Theorem 8.1 requires something stronger than Theorem 8.10; it requires a version of this for measures which is given in Theorem 8.14. This can be a subtle point.

As far as the second fundamental theorem of calculus, we will stick to 1 dimension but try to understand exactly when it holds. Recall for the Cantor ternary function, it failed. The answer to when it holds will be intimately connected to the concept of absolute continuity which we have encountered in the previous chapter. Unlike the first fundamental theorem of calculus, we will not look for a generalization of the second fundamental theorem of calculus in higher dimensions. In nice cases, this would correspond to Stokes Theorem in vector calculus.

8.3 Approximations by sets and by continuous functions

The following discussion appeared much earlier in F but we have postponed it since we didn't need it until now. We will state the results clearly but we will not cover the proofs. They are not so difficult in any case. They are very related to exercise 18 in Chapter 1 of F which we presented in the exercises.

First some background: the Borel σ-algebra on R^n, denoted by $\mathcal{B}^n$, can be equivalently described as either (1) the σ-algebra generated by the open sets in R^n or (2) the product σ-algebra coming from $(R, \mathcal{B})$. Proposition 1.5 in F says that these are equivalent.

Theorem 8.3. *Let m be Lebesgue measure on R^n.*
1. (Theorem 2.40 in F) For any Borel set $E \subseteq R^n$ (and even for any Lebesgue measurable set), we have

$$m(E) = \sup\{m(K) : K \subseteq E, K \text{ compact }\}.$$

2. (Theorem 2.41 in F) If $f \in L^1(R^n, \mathcal{B}^n, m)$ and $\epsilon > 0$, then there exists a continuous function g on R^n so that

$$\int_{R^n} |f(x) - g(x)| dm(x) < \epsilon$$

We will also need to approximate sets with respect to other measures.

Definition 8.4. *A Borel measure ν on R^n (meaning a measure on $(R^n, \mathcal{B}^n)$) is* **regular** *if $\nu(K) < \infty$ for all compact sets $K \subseteq R^n$.*

Theorem 8.5. *Let ν be a regular Borel measure on R^n. If E is a Borel set, then*

$$\nu(E) = \inf\{\nu(O) : O \supseteq E, O \text{ open }\}.$$

It is the case that the statements in Theorem 8.3 are also true for all Borel measures on R^n but I wanted to just write down the minimum which we need.

8.4 A covering lemma

There are many types of so-called covering lemmas in analysis. The one which will be useful for our purposes here is the following. m will always denote Lebesgue measure.

Theorem 8.6. *(A covering lemma)*
Let $\mathcal{C}$ be a collection of open balls in R^n with U being their union. For all $c < m(U)$, there exist $B_1, B_2, \ldots, B_k \in \mathcal{C}$ which are disjoint satisfying

$$\sum_{i=1}^{k} m(B_i) \geq \frac{c}{3^n}.$$

Note that in the special case that $\mathcal{C}$ is finite, this says that we can cover a fixed uniform proportion of U by disjoint balls from $\mathcal{C}$.

Proof:
By Theorem 8.3(1), choose $K \subseteq U$ compact so that $m(K) > c$ and by compactness, choose $A_1, \ldots, A_m \in \mathcal{C}$ which cover K. Now, restricting to $A_1, \ldots, A_m$, we choose B_1 among these to have maximal size. Then choose B_2 among these to have maximal size but disjoint from B_1. Then choose B_3 among these to have maximal size but disjoint from $B_1 \cup B_2$. Continue as far as possible. Let $B_1, B_2, \ldots, B_k$ be the obtained sets which are certainly disjoint.

Claim: If A_i is not in the final list $B_1, B_2, \ldots, B_k$, then for some j

$$A_i \subseteq B_j^* \tag{18}$$

where B_j^* is the ball concentric with B_j but with three times the radius. To see this, since A_i was not chosen, it must have intersected one of the B_j's. The first B_j which A_i intersects must be at least as large as A_i since otherwise we would have chosen A_i instead of B_j at that stage. (18) now follows geometrically.

(18) implies that $\bigcup_{j=1}^k B_j^* \supseteq K$ which in turn yields

$$c < m(K) \leq m(\bigcup_{j=1}^{k} B_j^*) \leq \sum_{j=1}^{k} m(B_j^*) = 3^n \sum_{j=1}^{k} m(B_j).$$

QED

8.5 The Hardy-Littlewood Maximal Theorem in Euclidean Space

$B(x,r)$ will denote the open ball about x with radius r.

Definition 8.7. *A function $f : R^n \to R$ is said to be in L^1_{loc} if f is integrable with respect to Lebesgue measure when restricted to any compact set (equivalently, if for every x, there is an r so that f is integrable on $B(x,r)$ with respect to Lebesgue measure).*

For $f \in L^1_{loc}$, $x \in R^n$ and $r > 0$, we define

$$A_r f(x) := \frac{1}{m(B(x,r))} \int_{B(x,r)} f(y)dm(y)$$

which is simply the average of f over the ball around x with radius r.

Remarks:
1. If f is continuous at x, it is easy to see (check this) that

$$\lim_{r\to 0} A_r f(x) = f(x).$$

2. It is clear that this cannot occur for all f and for all x since if we modify a continuous function by changing its value at one point, this convergence would clearly fail at that point.
3. Perhaps one would hope that the above convergence would occur at a.e. x for any f; this is indeed the case and one of the things that we want to prove.
4. It is believable and not so hard to prove that for any $f \in L^1_{loc}$, $A_r f(x)$ is a continuous function of (r,x); we will not prove this, see Lemma 3.16 in F.

The following so-called Hardy-Littlewood Maximal function and its subsequent theorem will be crucial to prove our first "differentiation theorem".

Definition 8.8. *If $f \in L^1_{loc}(R^n)$, then the* **Hardy-Littlewood Maximal function** *Hf is defined by*

$$Hf(x) := \sup_{r>0} A_r|f|(x).$$

Exercise: If $f \in L^1(R^n)$, then for all x one has $\lim_{r\to\infty} A_r|f|(x) = 0$.

One checks that Hf is a measurable function by noting that

$$(Hf)^{-1}(\alpha,\infty) = \bigcup_{r>0} (A_r|f|)^{-1}(\alpha,\infty)$$

is open by remark 4 above.

Theorem 8.9. *(Hardy-Littlewood Maximal Theorem)*
If $f \in L^1(R^n)$, then for all $\alpha > 0$,

$$m(\{x : Hf(x) > \alpha\}) \leq \frac{3^n}{\alpha} \int |f| dm(x)$$

Proof:
Let $E_\alpha := \{x : Hf(x) > \alpha\}$. For each $x \in E_\alpha$, there exists $r_x > 0$ so that

$$A_{r_x}|f|(x) > \alpha.$$

Given $c < m(\bigcup_{x \in E_\alpha} B(x, r_x))$, Theorem 8.6 allows us to find $x_1, \ldots, x_k$ in E_α so that the k balls $\{B(x_i, r_{x_i})\}$ are disjoint and

$$\frac{c}{3^n} \leq \sum_{i=1}^{k} m(B(x_i, r_{x_i})).$$

However, by the way we have chosen the r_x's, we have

$$\sum_{i=1}^{k} m(B(x_i, r_{x_i})) \leq \sum_{i=1}^{k} \frac{1}{\alpha} \int_{B(x_i, r_{x_i})} |f|(x) dm(x) \leq \frac{1}{\alpha} \int |f|(x) dm(x)$$

where the last inequality follows from the disjointness of the balls. Since this is true for all $c < m(E_\alpha)$, we are done.
QED

8.6 Differentiation of functions with respect to Lebesgue measure in R^n

Theorem 8.10. *(The differentiation theorem for functions)*
If $f \in L^1_{loc}(R^n)$, then for Lebesgue a.e. point x

$$\lim_{r \to 0} A_r f(x) = f(x).$$

Remarks before starting the proof.
(i) The proof uses the notion of a $\limsup$ for functions, not just sequences; this is done in the "obvious" way, see F, pg 96 if not clear.
(ii) In the proof, we will introduce a set called E_α; this is not the same E_α which was defined in the previous proof. However, since these sets are only "locally defined" in a proof, no confusion should arise.

Proof:
It suffices to show this for a.e. point in $[-N, N]^n$ and therefore we may assume f is 0 outside of $[-N-1, N+1]^n$ in which case we have that $f \in L^1(R^n)$. Fix α and let

$$E_\alpha := \{x : \limsup_{r\to 0} |A_r f(x) - f(x)| > \alpha\}.$$

If we can show that $m(E_\alpha) = 0$ for each α, then this implies the theorem since the set of points where the conclusion of the theorem fails is exactly equal to $\bigcup_n E_{\frac{1}{n}}$ which then also has measure 0. (Note that it is not apriori obvious that E_α is Lebesgue measurable; however the proof will show that it is contained inside measurable sets of arbitrarily small measure. It will then follow that its outer measure is zero and hence by completeness it will be Lebesgue measurable. The same comment applies to the rest of the chapter.)

We now fix $\alpha > 0$ and show $m(E_\alpha) = 0$. Fix $\epsilon > 0$. By Theorem 8.3, we can find a continuous integrable function g with

$$\int_{R^n} |f(x) - g(x)| dm(x) < \epsilon.$$

Observe that

$$\limsup_{r\to 0} |A_r f(x) - f(x)| = \limsup_{r\to 0} |A_r f(x) - A_r g(x) + A_r g(x) - g(x) + g(x) - f(x)| =$$

$$\limsup_{r\to 0} |A_r(f-g)(x) + A_r g(x) - g(x) + g(x) - f(x)| \le$$

$$\limsup_{r\to 0} |A_r(f-g)(x)| + \limsup_{r\to 0} |A_r g(x) - g(x)| + \limsup_{r\to 0} |g(x) - f(x)|.$$

The first term is at most $H(f-g)(x)$ and the second term is 0 since the theorem is easily checked for continuous functions (details in F if needed). Hence

$$\limsup_{r\to 0} |A_r f(x) - f(x)| \le H(f-g)(x) + |g(x) - f(x)|.$$

If the first term is larger than α, then at least one of the terms on the RHS have to be larger than $\alpha/2$, i.e.,

$$E_\alpha \subseteq \{x : H(f-g)(x) > \frac{\alpha}{2}\} \cup \{x : |g(x) - f(x)| > \frac{\alpha}{2}\}.$$

Hence

$$m(E_\alpha) \le m(\{x : H(f-g)(x) > \frac{\alpha}{2}\}) + m(\{x : |g(x) - f(x)| > \frac{\alpha}{2}\}).$$

Using Theorem 8.9 for the first term and Markov's inequality on the second term, we get

$$m(E_\alpha) \le \frac{2(3^n)}{\alpha}\epsilon + \frac{2}{\alpha}\epsilon.$$

Since this is true for every $\epsilon > 0$, we can conclude $m(E_\alpha) = 0$.
QED

8.7 The Lebesgue set of a function and other "small sets shrinking nicely"

Theorem 8.10 which states that for any function f which is locally integrable, one has for Lebesgue a.e. point x

$$\lim_{r\to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} (f(y) - f(x))dm(y) = 0$$

this might be due to some cancellation. The following stronger result shows that this is not the case.

Theorem 8.11. *(Lebesgue points)*
For any function f which is locally integrable, one has for Lebesgue a.e. point x

$$\lim_{r\to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|dm(y) = 0. \tag{19}$$

(The set of x where this holds is called the **Lebesgue set** *of f and is denoted by L_f.)*

Proof:
For each $q \in Q$, let $g_q(x) := |f(x) - q|$. By Theorem 8.10 applied to g_q, we have

$$\lim_{r\to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - q|dm(y) = |f(x) - q|$$

for all x except a set E_q of Lebesgue measure 0. Since $m(\bigcup_{q\in Q} E_q)) = 0$, it is enough to show that (19) holds for $x \notin \bigcup_{q\in Q} E_q$. Fix such an x and let $q \in Q$. We have

$$\limsup_{r\to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|dm(y) \le$$

$$\limsup_{r\to 0} \frac{1}{m(B(x,r))} \int_{B(x,r)} (|f(y) - q| + |f(x) - q|)dm(y) \le$$

$$\limsup_{r\to 0} \frac{1}{m(B(x,r))}\int_{B(x,r)} |f(y)-q|dm(y) + |f(x)-q| =$$
$$|f(x)-q| + |f(x)-q|.$$

Since this inequality holds for all $q \in Q$, the LHS is 0.
QED

Having this stronger version allows us to replace the balls $\{B(x,r)\}$ by other types of sets. This will be useful when we prove Lebesgue's Theorem for monotone functions in one dimension, Theorem 9.1.

Definition 8.12. *A collection of Borel sets $\{E_r\}_{0<r\le 1}$ is said to* **shrink nicely to** *x if for all $r \in (0,1]$, $E_r \subseteq B(x,r)$ and there exists $\alpha > 0$ so that $m(E_r) \ge \alpha m(B(x,r))$ for all r. (The second condition says that each E_r occupies a substantial proportion of $B(x,r)$; also there is no requirement that $x \in E_r$.)*

Exercise: Show that in R^2, $E_r := [0, \frac{r}{2}] \times [0, \frac{r}{200}]$ shrinks nicely to $(0,0)$ but $E_r := [0, \frac{r}{2}] \times [0, \frac{r^2}{2}]$ does not shrink nicely to $(0,0)$.

Corollary 8.13. *Let f be locally integrable. If x belongs to the Lebesgue set L_f (and hence for m a.e. x) and $\{E_r\}_{0<r\le 1}$ are Borel sets shrinking nicely to x, then*

$$\lim_{r\to 0} \frac{1}{m(E_r))}\int_{E_r} |f(y)-f(x)|dm(y) = 0 \text{ and } \lim_{r\to 0} \frac{1}{m(E_r))}\int_{E_r} f(y)dm(y) = f(x). \tag{20}$$

Proof:
The second statement follows from the first and the first follows immediately from observing

$$\frac{1}{m(E_r))}\int_{E_r} |f(y)-f(x)|dm(y) \le \frac{1}{\alpha m(B(x,r))}\int_{B(x,r)} |f(y)-f(x)|dm(y).$$

QED

As a special case, if $f \in L^1([0,1])$, letting $E_r(x) := [x, x+r]$, one has that

$$\left(\int_0^x f(t)dm(t)\right)'(x) = f(x) \text{ for a.e. } x.$$

8.8 Differentiation of measures with respect to Lebesgue measure in R^n

Theorem 8.14. *(The differentiation theorem for measures)*
Let ν be a regular Borel measure on R^n whose Lebesgue decomposition with respect to Lebesgue measure is given by $\nu_s + fm$. Then

$$\lim_{r\to 0} \frac{\nu(B(x,r))}{m(B(x,r))} = f(x) \text{ exists for Lebesgue a.e. } x.$$

Remark: If $\nu \ll m$, this reduces to Theorem 8.10.

Proof:
By writing

$$\nu(B(x,r)) = \nu_s(B(x,r)) + \int_{B(x,r)} f(y)dm(y),$$

it suffices, in view of Theorem 8.10, to prove that for any regular Borel measure λ singular with respect to Lebesgue measure we have

$$\lim_{r\to 0} \frac{\lambda(B(x,r))}{m(B(x,r))} = 0 \text{ for Lebesgue a.e. } x. \tag{21}$$

By mutual singularity, there is a Borel set A with $\lambda(A) = 0 = m(A^c)$. By countable additivity, it is enough to show that for each ℓ

$$F_\ell := \{x \in A : \limsup_{r\to 0} \frac{\lambda(B(x,r))}{m(B(x,r))} > \frac{1}{\ell}\}$$

has Lebesgue measure 0.

Fix now $\epsilon > 0$ and by Theorem 8.5, choose an open set $U_\epsilon \supseteq A$ with $\lambda(U_\epsilon) < \epsilon$. For each $x \in F_\ell$ ($\subseteq A \subseteq U_\epsilon$), choose $r_x > 0$ such that $B(x, r_x) \subseteq U_\epsilon$ and

$$\lambda(B(x,r_x)) > \frac{m(B(x,r_x))}{\ell}.$$

Given $c < m(\bigcup_{x\in F_\ell} B(x,r_x))$, Theorem 8.6 allows us to find $x_1, \ldots, x_k$ in F_ℓ so that the k balls $B(x_i, r_{x_i})$ are disjoint and

$$\frac{c}{3^n} \le \sum_{i=1}^{k} m(B(x_i, r_{x_i})).$$

However, by the way we have chosen these balls, we have

$$\sum_{i=1}^{k} m(B(x_i, r_{x_i})) \leq \sum_{i=1}^{k} \ell\lambda(B(x_i, r_{x_i})) \leq \ell\lambda(U_\epsilon) < \ell\epsilon$$

where the second to last inequality follows from the disjointness of the balls. Hence $c < 3^n\ell\epsilon$ for all $c < m(\bigcup_{x\in F_\ell} B(x, r_x))$ and so

$$m(\bigcup_{x\in F_\ell} B(x, r_x)) \leq 3^n\ell\epsilon.$$

Finally, since $F_\ell \subseteq \bigcup_{x\in F_\ell} B(x, r_x)$, we have that

$$m(F_\ell) \leq 3^n\ell\epsilon$$

for all $\epsilon > 0$. This gives $m(F_\ell) = 0$ as desired.
QED

Remark: Sometimes it is important for understanding to take a step back and try to see what is happening at a more heuristic level. The key part is to show that $m(F_\ell) = 0$. Why should this be true? At each point of F_ℓ, λ is giving small balls measure at least $\frac{1}{\ell}$ times their Lebesgue measure. In other words, near each point of F_ℓ, λ is "expanding" things by a factor of at least $\frac{1}{\ell}$. Since this is happening at every point of F_ℓ, it is reasonable that

$$\lambda(F_\ell) \geq \frac{1}{\ell} m(F_\ell). \tag{22}$$

If one had this, one is done since $F_\ell \subseteq A$ implying the LHS is 0 and hence $m(F_\ell) = 0$. Finally, the point of the covering lemma is to allow us to rigorously go from the "local information" that we have to "global information" given by some weaker variant of (22).

In the exact same way that Corollary 8.13 was proved, one can obtain the following corollary from Theorem 8.14.

Corollary 8.15. *Let ν be a regular Borel measure on R^n whose Lebesgue decomposition with respect to Lebesgue measure is given by $\nu_s + fm$. Then for Lebesgue a.e. x, we have that if $\{E_r\}_{0<r\leq 1}$ are Borel sets shrinking nicely to x, then*

$$\lim_{r\to\infty} \frac{\nu(E_r)}{m(E_r)} = f(x).$$

8.9 What is happening on the singular set itself? An interesting aside

In this section, we give a result complementary to our last result. More specifically, (21) tells us that for a singular measure, the limit of the ratios of the measures of the balls is 0 for a typical point with respect to Lebesgue measure. One can now ask how the limit of the ratios of the measures of the balls behave for a typical point with respect to the singular measure λ instead. The next result tells us this is ∞ and the proof is similar to obtaining (21).

Theorem 8.16. *(The differentiation theorem on a singular set)*
Let λ be a regular Borel measure on R^n singular with respect to Lebesgue measure. Then

$$\lim_{r\to 0}\frac{\lambda(B(x,r))}{m(B(x,r))}=\infty \text{ for } \lambda \text{ a.e. } x.$$

Proof:
By mutual singularity, there is a Borel set S with $\lambda(S^c)=0=m(S)$. By countable additivity, it is enough to show that for each ℓ

$$F_\ell := \{x\in S : \liminf_{r\to 0}\frac{\lambda(B(x,r))}{m(B(x,r))}<\ell\}$$

has λ measure 0. It suffices by Theorem 8.3(1) (using the comment later that this result applies to all Borel measures), to show that for all compact $C\subseteq F_\ell$, $\lambda(C)=0$.

Fix now $\epsilon>0$ and by Theorem 8.5 (or directly from the definition of Lebesgue measure), choose an open set $U_\epsilon\supseteq S$ with $m(U_\epsilon)<\epsilon$. For each $x\in C\,(\subseteq F_\ell\subseteq S\subseteq U_\epsilon)$, choose $r_x>0$ such that $B(x,r_x)\subseteq U_\epsilon$ and

$$\lambda(B(x,r_x))\le \ell m(B(x,r_x)).$$

By compactness, there exist $x_1,\ldots,x_n\in C$ so that $C\subseteq\bigcup_{i=1}^n B(x_i,\frac{r_{x_i}}{3})$. By the proof of Theorem 8.6, one can find a subcollection of these balls, corresponding to $y_1,\ldots,y_m\in C$, so that these balls are disjoint and $C\subseteq\bigcup_{j=1}^m B(y_j,r_{y_j})$.

We now have, using disjointness for the last equality,

$$\lambda(C)\le\lambda(\bigcup_{j=1}^m B(y_j,r_{y_j}))\le\sum_{j=1}^m\lambda(B(y_j,r_{y_j}))\le\ell\sum_{j=1}^m m(B(y_j,r_{y_j}))$$

$$=\ell 3^n\sum_{j=1}^m m(B(y_j,\frac{r_{y_j}}{3}))=\ell 3^n m(\bigcup_{j=1}^m B(y_j,\frac{r_{y_j}}{3}))\le \ell 3^n m(U_\epsilon)\le \ell 3^n\epsilon.$$

QED

8.10 An alternative proof of the Differentiation of functions with respect to Lebesgue measure in R^n avoiding the Hardy-Littlewood Maximal inequality and approximation by continuous functions: An interesting aside

There are other possible approaches to the differentiation theorems than those which were presented here. Some can avoid both the Hardy-Littlewood Maximal inequality and approximation by continuous functions. Here we follow an approach by Rudin (Real and Complex Analysis, second but not third edition).

The first step is in fact part of the proof of Theorem 8.14; this will not be a circular argument.

Lemma 8.17. *Let λ be a Borel measure on R^n and let A be a Borel set with $\lambda(A) = 0$. Then there exists $A' \subseteq A$ so that*

$$(i)\ m(A \backslash A') = 0$$

and

$$(ii)\ \lim_{\epsilon \to 0} \frac{\lambda(B(x,\epsilon))}{m(B(x,\epsilon))} = 0 \text{ for } x \in A'.$$

Proof:
The proof of this is contained in the second half of the proof of Theorem 8.14, starting from (21). Looking at this proof, it shows exactly that for any A with $\lambda(A) = 0$ (not assuming any mutual singularity), we obtain A' as claimed in the lemma, namely $A' = A \backslash \left(\bigcup_{\ell=1} F_\ell\right)$.
QED
Note that this uses Theorems 8.5 and 8.6 but avoids the Hardy-Littlewood Maximal inequality and approximation by continuous functions.

Alternative Proof of Theorem 8.10 avoiding HLM and approximation:
Let $\lambda := f dm$ so that f is the Radon-Nikodym derivative of λ with respect to Lebesgue measure. We need to show that

$$\lim_{\epsilon \to 0} \frac{\lambda(B(x,\epsilon))}{m(B(x,\epsilon))} = f(x) \text{ for } m - a.e.\ x$$

Fix $r \in Q$ and let $A_r := \{x : f(x) < r\}$ and $B_r := \{x : f(x) \geq r\} (= A_r^c)$. Define the Borel measure λ_r by

$$\lambda_r(E) := \int_{E \cap B_r} (f(x) - r) dm(x).$$

In words, λ_r is the measure concentrated on B_r with Radon-Nikodym derivative $(f(x)-r)I_{B_r}$ with respect to Lebesgue measure. Note that λ_r is a positive measure. Since $\lambda_r(A_r) = 0$, by the previous lemma, there exists $A'_r \subseteq A_r$ with

$$m(A_r \backslash A'_r) = 0$$

and

$$\lim_{\epsilon \to 0} \frac{\lambda_r(B(x,\epsilon))}{m(B(x,\epsilon))} = 0 \text{ for } x \in A'_r. \tag{23}$$

Letting $Y := \bigcup_{r \in Q} (A_r \backslash A'_r)$, we have $m(Y) = 0$. It now suffices to show that

$$\lim_{\epsilon \to 0} \frac{\lambda(B(x,\epsilon))}{m(B(x,\epsilon))} = f(x) \text{ for } x \notin Y.$$

Fix $x \notin Y$ and and choose $r > f(x)$ with $r \in Q$. Note that for any set E

$$\lambda(E) - rm(E) = \int_E (f(y) - r) dm(y) \leq \lambda_r(E)$$

the inequality holding since $\lambda_r(E)$ is a restriction of the integral to where the integrand is nonnegative. This yields

$$\frac{\lambda(B(x,\epsilon))}{m(B(x,\epsilon))} \leq \frac{\lambda_r(B(x,\epsilon))}{m(B(x,\epsilon))} + r.$$

Since $x \in A_r \backslash Y$, we have $x \in A'_r$ and hence, using (23), we obtain

$$\limsup_{\epsilon \to 0} \frac{\lambda(B(x,\epsilon))}{m(B(x,\epsilon))} \leq r.$$

Since this is true for all $r > f(x)$ with $r \in Q$, we obtain

$$\limsup_{\epsilon \to 0} \frac{\lambda(B(x,\epsilon))}{m(B(x,\epsilon))} \leq f(x).$$

So this inequality holds for m-a.e. x.
Applying the same argument to $-f$ which corresponds to $-\lambda$, we have

$$\limsup_{\epsilon \to 0} \frac{-\lambda(B(x,\epsilon))}{m(B(x,\epsilon))} \leq -f(x)$$

or

$$\liminf_{\epsilon \to 0} \frac{\lambda(B(x,\epsilon))}{m(B(x,\epsilon))} \geq f(x)$$

for m-a.e. x.
QED

9 Differentiation in 1-dimension: Lebesgue's famous Theorem, Bounded variation, absolute continuity

We begin this section with proving Lebesgue's Theorem.

9.1 Lebesgue's differentiation theorem in 1-dimension

Theorem 9.1. *Let F be a right continuous increasing function on $[0, 1]$. Then F has a finite derivative at Lebesgue a.e. x.*

Remarks:
(1). The theorem is true without the right continuity assumption (see Theorem 3.23 in F.); however this assumption makes things simpler.
(2). Before starting the proof, we point out that there is an issue here which can be confusing. Even though we are looking at differentiation of the particular function F, it is not Theorem 8.10 and Corollary 8.13 which will be used but rather Theorem 8.14 and Corollary 8.15.

Proof:
Let ν be the Borel measure on $[0, 1]$ associated with F so that

$$\nu((a, b]) = F(b) - F(a).$$

Let $\nu = \nu_s + f dm$ be the Lebesgue decomposition of ν with respect to m. We have that for $h > 0$

$$\frac{F(x+h) - F(x)}{h} = \frac{\nu((x, x+h])}{m((x, x+h])}$$

and so Corollary 8.15 (with $E_r = (x, x+r]$) tells us that

$$\lim_{h \downarrow 0} \frac{F(x+h) - F(x)}{h} = f(x)$$

for Lebesgue a.e. x. Next, for $h < 0$

$$\frac{F(x+h) - F(x)}{h} = \frac{\nu((x+h, x])}{m((x+h, x])}$$

and so Corollary 8.15 (with $E_r = (x-r, x]$) tells us that

$$\lim_{h \uparrow 0} \frac{F(x+h) - F(x)}{h} = f(x)$$

for Lebesgue a.e. x.
QED

Remark: The proof shows that the derivative of F corresponds a.e. to the Radon-Nikodym derivative of the absolute continuous part of ν with respect to m. This corresponds to the fact in elementary probability that, in nice cases, the probability density function is simply the derivative of the distribution function.

Remark: The following fact is not covered in most books. It turns out that if F is monotone increasing on $[0,1]$ and is discontinuous at a dense set of points, then, although F is differentiable a.e., it is necessarily the case that the set of points where F is **not** differentiable is very large topologically in that it contains a dense G_δ (see Subsection 3.14).

9.2 Bounded variation and absolute continuity

We now begin to study some important properties of functions from $[0,1]$ to R. We will also be discussing the relationship between these properties and properties of the associated measures in the sense of Subsection 3.10.

Definition 9.2. *If $f:[0,1]\to R$, we define the* **total variation** *of f on $[a,b]\subseteq[a,b]$ to be*

$$TV_{[a,b]}(f) := \sup\{\sum_{i=1}^{n} |f(x_i)-f(x_{i-1})| : a = x_0 < x_1 < x_2 < \ldots < x_n = b\}. \tag{24}$$

We say f **is of bounded variation** *on $[a,b]$ if $TV_{[a,b]}(f)<\infty$; otherwise we say f* **is of unbounded variation** *on $[a,b]$.*

Remarks:
(i). If f is monotone increasing, then $TV_{[a,b]}(f) = f(b)-f(a)$.
(ii). $TV_{[a,b]}(-f) = TV_{[a,b]}(f)$.
(iii). If f is the indicator function of the rationals, then f is of unbounded variation on every (nontrivial) interval.

Exercises:
1. Check all the remarks above.
2. If f has a continuous derivative on $[a,b]$, then f is of bounded variation on $[a,b]$.
3. If f is Lipschitz continuous on $[a,b]$, then f is of bounded variation on

$[a, b]$.
4. Show that the continuous function

$$f(x) = x\cos(\frac{\pi}{2x}) \text{ (defined to be 0 at 0)}$$

is of unbounded variation on $[0, 1]$ but of bounded variation on $[\epsilon, 1]$ for each $\epsilon > 0$. (So the "problem" is occuring at 0.)

The following is the key characterization of functions of bounded variation.

Theorem 9.3. *f is of bounded variation on an interval if and only if there exist increasing functions g and h so that $f = g - h$.*

Proof Outline:

The "if" direction is quite easy and left as an exercise. The "only if" direction is more difficult and more interesting.

Step 1: Claim:

$$TV_{[a,c]}(f) = TV_{[a,b]}(f) + TV_{[b,c]}(f).$$

Every partition of $[a, b]$ and $[b, c]$ yields a partition of $[a, c]$ which easily yields $\geq$. Next, using the fact that the sums in (24) can only increase with a finer partition, we can assume all partitions of $[a, c]$ which we consider include b which then easily yields $\leq$. In particular $TV_{[a,x]}(f)$ is an increasing function of x.

Step 2: $TV_{[0,x]}(f) + f(x)$ is an increasing function of x.
subproof: If $0 \leq x < y \leq 1$,

$$f(x) - f(y) \leq |f(x) - f(y)| \leq TV_{[x,y]}(f) = TV_{[0,y]}(f) - TV_{[0,x]}(f)$$

where the last equality comes from Step 1. Now rewrite. subQED

Step 3: Note that

$$f(x) = \frac{TV_{[0,x]}(f) + f(x)}{2} - \frac{TV_{[0,x]}(f) - f(x)}{2}.$$

The two summands are increasing in x by Step 2, where for the second term we also use the fact that $TV_{[0,x]}(-f) = TV_{[0,x]}(f)$.
QED

There is a 1-1 correspondence between signed measures and right-continuous functions of bounded variation. The bijection is given by μ a signed measure on $[0,1]$ is sent to the bounded variation function

$$F_\mu(x) := \mu([0,x].$$

This is elaborated in the following discussion.

Exercise: Prove the claim above.

Discussion: If f is of bounded variation, then one can define the so-called Lebesgue-Stieltjes integral of a function r with respect to f which is defined as

$$\int r(x)df(x) := \int r(x)d\mu_g(x) - \int r(x)d\mu_h(x)$$

where g, h are increasing, $f = g-h$ and μ_g and μ_h are the measures associated to g and h. If f is not of bounded variation, then $\int r(x)df(x)$ is not defined. In nice cases, for example if f is continuously differentiable, then it turns out that

$$\int r(x)df(x) = \int r(x)f'(x)dx. \tag{25}$$

In addition, by Theorems 9.1 and 9.3 when f is of bounded variation, f' exists and is finite a.e. Therefore the RHS of (25) is then defined but nonetheless (25) may fail. (This is similar to the failure of the fundamental theorem of calculus for the Cantor function defined earlier.) To understand when (25) will hold brings us to our next important concept.

Definition 9.4. *$f : [0,1] \to R$ is* **absolutely continuous** *if for all $\epsilon > 0$, there exists $\delta > 0$ so that if $0 \le x_1 < y_1 < x_2 < y_2 < \ldots, x_n < y_n \le 1$ and*

$$\sum_{i=1}^{n}(y_i - x_i) < \delta,$$

then

$$\sum_{i=1}^{n} |f(y_i) - f(x_i)| < \epsilon.$$

Clearly absolute continuity implies continuity and hence uniformity continuity since we are in a compact interval.

Exercise: Show that the Cantor ternary function, which is continuous, is not absolutely continuous.

The expression "absolute continuity" has been used in two completely different contexts. The next proposition justifies this terminology.

Proposition 9.5. *(Proposition 3.32 in F.) Let f be a nonnegative right-continuous monotone increasing function on $[0,1]$ with $f(0)=0$. Then f is absolutely continuous if and only if $\mu_f \ll m$ where m is Lebesgue measure.*

Proof:
First, note that f is continuous if and only if μ_f has no atoms. If these equivalent conditions fail, then both sides in the proposition fail. Hence we can assume that f is continuous or equivalently μ_f is continuous (i.e. no atoms).

Now, the "if" direction is fairly straightforward. Let $\epsilon > 0$ and choose $\delta > 0$ using Proposition 7.10 so that

$$m(A) < \delta \text{ implies that } \mu_f(A) < \epsilon.$$

Now if $0 \leq x_1 < y_1 < x_2 < y_2 < \ldots, x_n < y_n \leq 1$ with $\sum_{i=1}^n (y_i - x_i) < \delta$, then we have that

$$m(\bigcup_{i=1}^n (x_i, y_i)) < \delta$$

implying that

$$\mu_f(\bigcup_{i=1}^n (x_i, y_i)) < \epsilon$$

which is equivalent to

$$\sum_{i=1}^n |f(y_i) - f(x_i)| < \epsilon.$$

To prove the "only if" direction, assume that $m(A) = 0$ for some Borel set A. We need to show that $\mu_f(A) = 0$. Fix $\epsilon > 0$ and choose the corresponding δ in the definition of absolute continuity of f. Let U be an open set containing A with $m(U) < \delta$ and write U as a disjoint union of open intervals $\{(a_i, b_i)\}$. Since we have for any N

$$\sum_{i=1}^N (b_i - a_i) < \delta,$$

it follows that

$$\sum_{i=1}^N |f(b_i) - f(a_i)| < \epsilon$$

and so $\mu_f(\bigcup_{i=1}^N (a_i, b_i)) < \epsilon$. By letting $N \to \infty$, we have $\mu_f(U) \leq \epsilon$. Since $A \subseteq U$, this gives $\mu_f(A) \leq \epsilon$ and since ϵ is arbitrary, we get $\mu_f(A) = 0$, as

desired.
QED

The following easy result relates Bounded Variation and absolute continuity.

Proposition 9.6. *If $f : [0,1] \to R$ is absolutely continuous, then it has bounded variation.*

Proof:
Let δ correspond to $\epsilon = 1$ in the definition of absolute continuity for f. Choose N to be an integer larger than $1/\delta$. Choose an arbitrary partition $0 = x_0 < x_1 < x_2 < \ldots < x_n = 1$. Since refining a partition only increases the sum in the definition of total variation, we can assume that $x_0 < x_1 < x_2 < \ldots < x_n$ contain the points k/N for each integer k. Then by breaking

$$\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|$$

into pieces corresponding to $[0, 1/N], [1/N, 2/N], \ldots, [(N-1)/N, 1]$, the sum over each piece is at most $\epsilon = 1$ since the length of each interval is less than δ. Since there are N intervals, we get a bound of N on the total variation.
QED

Remark: The Cantor Ternary function has bounded variation being increasing, showing that bounded variation does not imply absolute continuity.

9.3 The fundamental theorem of calculus

Recall that an increasing function has a derivative a.e. One can ask if the fundamental theorem of calculus holds.
Question: If $f : [0,1] \to R$ is increasing, does

$$\int_0^1 f'(x) = f(1) - f(0)?$$

Note that for the Cantor Ternary function, the LHS is 0 and the RHS is 1. This is indicative of how this inequality may fail for monotone increasing functions.

Proposition 9.7. *If $f : [0,1] \to R$ is monotone increasing, then*

$$\int_0^1 f'(x) \leq f(1) - f(0).$$

We will see later why this is true by interpreting everything in terms of measures. Here we give a fairly easy direct proof based upon Fatou's Lemma.

Proof:
Given any increasing f and $h \in (0,1)$, we let, for $x \in [0,1]$,

$$\mathrm{Diff}_h f(x) := \frac{f(x+h) - f(x)}{h}$$

be a discrete approximation to the derivative. In order that $\mathrm{Diff}_h f(x)$ is defined for all x, we extend f to have value $f(1)$ on $[1, 1+h]$. (Since h will go to 0, this will be of no significance.)

Next let, for $h \in (0,1)$,

$$\mathrm{Av}_h f(x) := \frac{1}{h} \int_x^{x+h} f(t)dt$$

be the average of f on $[x, x+h]$. It is elementary to check that for each $h \in (0,1)$,

$$\int_0^1 \mathrm{Diff}_h f(x) = \mathrm{Av}_h f(1) - \mathrm{Av}_h f(0).$$

Now, we have that for a.e. x

$$\lim_{h \to 0} \mathrm{Diff}_h(x) = f'(x)$$

and all functions are nonnegative since f is increasing. It then follows from Fatou's Lemma that

$$\int_0^1 f'(x)dx \leq \liminf_{h \to 0} \int_0^1 \mathrm{Diff}_h(x)dx$$

and the RHS equals by the above

$$\liminf_{h \to 0} (\mathrm{Av}_h f(1) - \mathrm{Av}_h f(0)) \leq f(1) - f(0).$$

QED

The following theorem is a consequence of a number of the theorems we have previously proved.

Remark: To keep things simpler, we assumed that f is monotone. Things can also be done when f is of bounded variation in which case μ_f is defined but is a signed measure. The formulation of the result in that more general case is however a little bit different.

Theorem 9.8. *Let f be a nonnegative right-continuous monotone increasing function on $[0,1]$ with $f(0) = 0$. Let μ_f be the associated measure on $[0,1]$ and consider the Lebesgue decomposition of μ_f with respect to Lebesgue measure*

$$\mu_f = \mu_s + \mu_{ac}.$$

Then the following hold.
1. The Radon-Nikodym derivative of μ_{ac} with respect to Lebesgue measure is given by f'.
2.

$$\mu_{ac}[0,1] = \int_0^1 f'(x)dx.$$

3. f is absolutely continuous if and only if $\int_0^1 f'(x)dx = f(1) - f(0)$. (So the second fundamental theorem of calculus holds if and only if f is absolutely continuous.)
4. μ_f is singular if and only if $f'(x) = 0$ a.e.

Proof:
The first statement is stated in the remark right after Theorem 9.1. This immediately implies the second statement. For the third statement, using Proposition 7.10, we have that f is absolutely continuous if and only if μ_f is absolutely continuous with respect to Lebesgue measure if and only if $\mu_f = \mu_{ac}$ if and only if $\mu_f[0,1] = \mu_{ac}[0,1]$ (since we are dealing with measures and not signed measures) if and only if $f(1) - f(0) = \mu_{ac}[0,1]$ if and only if (step 2) $\int_0^1 f'(x)dx = f(1) - f(0)$. For the fourth statement, μ_f is singular if and only $\mu_{ac}[0,1] = 0$ if and only if (step 2) $\int_0^1 f'(x)dx = 0$ if and only if $f'(x) = 0$ a.e.
QED