Mathematics Notes

Jason Hall

2022-2023

1 Mathematical Logic

1.1 What is a proof?

The primitives of mathematical logic[1] are what are called propositions. For instance, let the proposition P be "It rained on June 27th, 2022 in Washington, DC." The key characteristic of propositions is that they have the capability of having a truth-value. For our purposes, we'll say that a proposition P has a truth-value if P is either true or false.[2].

Propositions are divided into two classes: simple propositions and compound propositions. To see the distinction, consider the following example. Let the proposition Q be "The author of these notes is Jason Hall." Both P and Q are simple propositions. Now consider the proposition given by "It rained on June 27th, 2022 in Washington, DC **and** the author of these notes is Jason Hall." This is a compound proposition because it is given by the conjunction of P and Q with the logical connective "and", which we typically denote by $\wedge$. Thus this proposition we would denote by $P \wedge Q$. As with a simple proposition, we can evaluate the truth-value of this compound proposition. To do so, however, we need to recognize that $\wedge$ is what is called a truth-functional. It takes as input the constituent truth values of the simple propositions that it connects, and returns a truth value for the compound statement as a whole.

As with any function, we need to know what maps to what. The way this has traditionally been done is with a truth table, essentially a list that says, given the inputs, what is the output. The truth table for $\wedge$ is given by

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

This means that our statement $P \wedge Q$ is true, given that both P and Q are true.[3]

In the traditional propositional calculus, there are 5 connectives:

[1] As a somewhat pedantic aside, what I am calling mathematical logic here is more properly termed the traditional propositional calculus. Mathematical logic typically refers to the field studying logical systems like the one described here.

[2] Technically, this is what is referred to as the principle of bivalence, which states that every statement is true or false. One does not need to accept the principle of bivalence, and one of the current research areas in logic looks at statements where truth is not necessarily a binary. See e.g. fuzzy logic for more on this.

[3] One reads a truth table by finding the row associated row of truth values on the left, and then reading the truth value in that row in the column associated with the compound proposition

1. $\wedge$: conjunction - this is analogous to the word "and"
2. $\vee$: disjunction - this is analogous to the word "or"
3. $\sim$: negation - this is analogous to the word "not"
4. $\implies$: implication - this is analogous to statements of the form "if ... then ..."
5. $\iff$: biconditional - this is analogous to statements of the form "... if and only if ..."

These five connectives give rise to compound propositions of the following form:

P	Q	$P \wedge Q$	$P \vee Q$	$\sim P$	$P \implies Q$	$P \iff Q$
T	T	T	T	F	T	T
T	F	F	T	F	F	F
F	T	F	T	T	T	F
F	F	F	F	T	T	T

I leave it as an exercise to the reader to show that $P \iff Q$ is logically equivalent to $(P \implies Q) \wedge (Q \implies P)$.

Returning back to the nature of statements, one should notice that the simple propositional logic above is going to have problems with statements of the form "Some of the apples are rotten". This is because we want to refer to subsets of an overarching object. To do this, we add quantifiers to the traditional propositional calculus. There are 3 quantifiers that are traditionally used:

1. Universal quantifier: $\forall$, this says that every element of an object x, has a property $P(x)$.
2. Existential quantifier: $\exists$, this says that at least one element of an object x has a property $P(x)$
3. Unique existential quantifier: $\exists!$, this says that exactly one element of an object has a property $P(x)$.[4]

How these quantifiers operate is governed by 4 rules:

1. Universal instantiation: Suppose that $\forall x \in X P(x)$ is true, then it follows that $c \in X$, $P(c)$ is true.
2. Universal generalization: Suppose that c has a property $P(c)$ for all c in the domain, then $\forall x P(x)$ is true.
3. Existential instantiation: Suppose that $\exists x \in X P(x)$ is true, then it follows that there is at least one $c \in X$, such that c has the property $P(c)$.
4. Existential generalization: Suppose that there is a c in the domain that has a property $P(c)$. Then it follows that $\exists x P(x)$ is true.

Now with those simple preliminaries out of the way, we now turn to the meat of what we are interested in, namely proving things. But to address that, we need to first answer two questions:

[4] This is technically redundant, one can write any case using $\exists!$, as a combination of $\forall$ and $\exists$. See for instance, $\exists! x \in \mathbb{N} | x = 1$. This can be equivalently written as $\exists x \in \mathbb{N} | x = 1 \wedge \sim \exists y \in \mathbb{N} | y = 1 \wedge y \neq x$.

1. What is an argument?

2. What is the relation to a proof?

The answer to these question is that an argument is a set of premises, and a set of conclusions. We say that an argument is valid if, given what we call rules of inference (which tell us what we can discover from a given premise), the conclusions follow from the premises. In classical propositional logic, there are 7 rules of inference.[5]

These are:

1. Addition
$$\begin{array}{rl} & P \\ \hline \therefore & P \vee Q \end{array}$$

2. Conjunction
$$\begin{array}{rl} & P \\ & Q \\ \hline \therefore & P \wedge Q \end{array}$$

3. Reduction
$$\begin{array}{rl} & P \wedge Q \\ \hline \therefore & P \end{array}$$

4. Modus Ponens
$$\begin{array}{rl} & P \implies Q \\ & P \\ \hline \therefore & Q \end{array}$$

5. Modus Tollens
$$\begin{array}{rl} & P \implies Q \\ & \sim Q \\ \hline \therefore & \sim P \end{array}$$

6. Disjunctive Syllogism
$$\begin{array}{rl} & P \vee Q \\ & \sim P \\ \hline \therefore & Q \end{array}$$

7. Hypothetical Syllogism
$$\begin{array}{rl} & P \implies Q \\ & Q \implies R \\ \hline \therefore & P \implies R \end{array}$$

The way to read the above is, given that the statements above the line are true, the statement below the line is true.

Now we are in position to answer both questions above. An argument is simply a set of premises and a conclusion. A valid argument is one where the conclusions follow from the premises, according to the rules of inference above. A proof is then an argument where we impose the requirement that our assumptions are true.

[5]Depending on the text you use, sometimes people will add 2 more.

1.2 Proving things

Traditionally, we divide proofs into five categories. Below I will go through each and provide an example.

- Direct Proof.

  Direct proofs are basically exactly as they sound. You are given a premise, and show that through some combination of the inference rules above, you directly recover the desired conclusion.

  For example, suppose one wanted to prove that the sum of two odd integers is even. To prove this, let x and y be odd numbers. Then x can be written as $2k_1 + 1$ for some $k_1 \in \mathbb{Z}$ and y can be written as $2k_2 + 1$ for some $k_2 \in \mathbb{Z}$. Then $x + y = 2k_1 + 1 + 2k_2 + 1 = 2k_1 + 2k_2 + 2 = 2(k_1 + k_2 + 1)$. Since $k_1 + k_2 + 1 \in \mathbb{Z}$, we thus have that $x + y$ is even, which was what was wanted.

- Proof by Contradiction

  The idea behind a proof by contradiction is simply to show that given a set of premises, if you make an assumption that is the opposite of what you want, you recover something that is inconsistent with your given premises.

  For example, suppose one wanted to prove the claim that there is no maximal natural number. To prove this, let us suppose that there is. Call this maximal natural number n. But $n + 1 \in \mathbb{N}$, by the closure of $\mathbb{N}$ under addition, and further $n + 1 > n$. Therefore it cannot be that n is maximal, which contradicts our assumption above. Hence we have the desired result.

- Proof by Exhaustion

  Suppose that you have an overarching set, that can be partitioned by some criterion.[6]. The game here is to consider all possible cases of the criterion, and examine where the desired proposition is true.

  For an example of proof by exhaustion, consider the following statement (you will go over this in Macro Mini 2, and you do actually need to know it!).

  Suppose that $W : \mathbb{R} \to \mathbb{R}$ is strictly concave[7] and differentiable, then it satisfies the following: $\forall x, y \in W$, $[W'(x) - W'(x')](x - x') \leq 0$, with equality if and only if $x' = x$.

  Our proof will proceed by considering three cases:

  1. $x < x'$: Suppose that $x < x'$. By the strict concavity of W, $W'(x) > W'(x')$. This implies that $[W'(x) - W'(x')] > 0$. Since $x < x'$, we have that $x - x' < 0$. Together, these imply that $[W'(x) - W'(x')](x - x') < 0$.
  2. $x > x'$: This case is symmetric with the prior case.
  3. $x = x'$: Suppose that $x = x'$. Again by strict concavity of W, $W'(x) = W'(x')$. This implies that $W'(x) - W'(x') = 0$, which further implies that $[W'(x) - W'(x')](x' - x) = 0$.

[6] For example, pick an element on the real line, call it r. Then $\forall x \in \mathbb{R}$ it follows that $x > r$, $x < r$, or $x = r$. This collection of possibilities is called the trichotomy.

[7] For our purposes, let's define a concave function f as one that has the property that its second derivative is less than zero.

Thus is the result shown.

- Proof by Contraposition.

  Proof by contraposition is simply an application of modus tollens above.

  Consider the following claim: if $x^2 - 4x + 1 \in \mathbb{Z}$ is even, then, x is odd.

  To prove this, let us suppose that x is even. Then x^2 is even.[8] Since x^2 is even, it can be written as $2k_2$ for some $k_2 \in \mathbb{Z}$. Thus $x^2 - 4x + 1 = 2k_2 - 2(2x) + 1 = 2(k_2 - 2x) + 1$. But since the integers are closed under both multiplication and addition, we have that $k_2 - 2x \in \mathbb{Z}$, which immediately gives us that if x is even, then $x^2 - 4x + 1$ is odd.

- Proof by Induction

  I will not cover this here, but rather in the next section where we cover basic set theory.

1.3 References

I highly recommend the book *How To Prove It* by Daniel Velleman. A free version can be found by clicking on the box at the end of this line.

[8]Suppose that x is even. Then $x = 2k$ for some $k \in \mathbb{Z}$. Then $x^2 = (2k)^2 = 4k^2 = 2(2k^2)$. Since $\mathbb{Z}$ is closed under multiplication, $2k^2 \in \mathbb{Z}$, and thus we have that x^2 is even.

2 Measure Theory

2.1 Introduction

One of the core concepts that we deal with in economics in general is that of uncertainty. When standing at a point in time, we do not know for sure what will happen on the next day, the next year, the next decade. To understand this uncertainty, we turn to probability, and the essence of probability relies upon the mathematical objects called measures. A measure is simply a generalization of the what people have traditionally called geometric measures: length, width, area, volume, etc.. The point of this section is to provide one with a basic background in the theory of measures. I highly highly recommend Terence Tao's free measure theory textbook, which you can find here. These notes are almost entirely based on my reading of this textbook (indeed they constitute the notes I took while working through the textbook). Some of the proofs are pulled exactly from there, while others are solutions to some of the exercises. Also as another note, I occasionally reference the cardinality of $\mathbb{R}$. For the sake of brevity[9], I assume that the continuum hypothesis is true. All mistakes are my own.

2.2 Motivation and the Problem of Measure

Modern mathematical rigor really only dates back to the late 19th century, and as a movement only really gained momentum in the early 20th century. Most of the initial research agenda was essentially in response to a set of questions about how one should measure subsets of the real line. This question, which is commonly called the problem of measure, is not as simple as it seems. To motivate general measure theory, it is helpful to see why exactly that is the case, and in the process learn a bit of history of mathematics. So let us begin by supposing that we have a set $A \subset \mathbb{R}^n$ which we would like to measure.

A natural starting point to measure is the so-called **elementary measure**. Elementary measure essentially adds up the sum of volumes of **boxes** - which are cartesian products of intervals, and have the usual volume (i.e. their volume is given by the product of the lengths of the intervals they are generated by)[10]. An elementary set is a finite union of such boxes. This has the advantage of conforming with our general intuitions about lengths and volumes, namely that they should be translation invariant, non-negative, and finitely additive if the boxes are disjoint. The disadvantage is, of course, that very few sets are measurable under this scheme. Spheres, triangular prisms, etc., are clearly not going to be measurable.

The response to this dilemma - indeed the natural generalization - comes from two different but clearly related objects, namely the **Jordan Outer Measure**:

$$m^{(J,*)}(A) = \inf_{A \subseteq O,\, O \in \xi(\mathbb{R}^n)} m(O)$$

and the **Jordan Inner Measure**:

$$m_{(J,*)}(A) = \sup_{I \subseteq A,\, I \in \xi(\mathbb{R}^n)} m(I)$$

We say that a set A in $\mathbb{R}^n$ is **Jordan Measurable** if its Jordan Outer Measure and Jordan Inner Measure coincide.

[9] I didn't want to type $2^{\aleph_0}$ every time.

[10] For convenience I will notate the space of elementary sets in $\mathbb{R}^n$ as $\xi(\mathbb{R}^n)$.

Intuitively, what this means, is that we can make approximations of some set both by enveloping it in boxes, and by constructing boxes inside it, and the measure of the these two approximations are arbitrarily close to each other.

This approach vastly increases the number of sets that can measure. All closed convex polytopes, for instance, are Jordan measurable. So are the graphs of continuous functions whose domain is a compact set.

Nevertheless, this is still not a sufficient way of measuring subsets of $\mathbb{R}^n$. For one thing, some subsets of $\mathbb{R}$ are still not measurable.

Claim: $\mathbb{Q} \cap [0,1]$ is not Jordan measurable.

Proof. Observe that the smallest elementary set that we can generate that will cover all of the relevant set is the interval $[0,1]$. If we had missed any segment of the interval, by the density of $\mathbb{Q}$ in $\mathbb{R}$, we would be able to pick a rational number in the missed segment, which would contradict the requirement that the elementary set contain all of the set we are trying to measure. Now observe that $\mathbb{Q} \cap [0,1]$ has empty interior, which immediately implies that the Jordan Inner Measure is 0. Hence $m^{(J,*)}(\mathbb{Q} \cap [0,1]) = 1 \neq 0 = m_{(J,*)}(\mathbb{Q} \cap [0,1])$. ■

More pressing, however, is the fact that the limit of a sequence of Jordan Measurable sets need not be Jordan measurable. The classical example of this is the so-called Fat Cantor Set. As much of analysis is concerned with the taking of limits, this is a sharp problem with Jordan measure.

As a result, Henri Lebesgue developed an alternative approach, which we today call the Lebesgue measure. The Lebesgue measure[11] of a set is built up from something called the **Lebesgue Outer Measure**:

$$\lambda^*(A) = \inf_{\cup_{i=1}^{\infty} B_i \supseteq A,\ B_i \text{boxes} \forall i} \sum_{i=1}^{\infty} |B_i|$$

The Lebesgue Outer Measure satisfies the **Outer Measure Axioms**:

1. (The empty set axiom) $\lambda^*(\emptyset) = 0$
2. (The monotonicity axiom) $A \subseteq B \subseteq \mathbb{R} \implies \lambda^*(A) \leq \lambda^*(B)$
3. (The countable subadditivity axiom) $\lambda^*(\cup_{i=1}^{\infty} B_i) \leq \sum_{i=1}^{\infty} \lambda^*(B_i)$

One can see Lebesgue Outer Measure in some ways as a natural generalization of Jordan Outer Measure, where we are going from the finite set of boxes required in the Jordan case, to a countable set of boxes in the Lebesgue measure. To see how these quantities relate, consider the following:

[11] Occasionally you see people refer to Lebesgue measurability as λ-measurability, which I am going to do from now on to preserve notation. To this end, the Lebesgue outer measure of a set will be notated $\lambda^*(\cdot)$, while the Lebesgue measure of a set will be denoted $\lambda(\cdot)$.

Claim: If E is an elementary set in $\mathbb{R}^n$, then $\lambda^*(E) = m^{(E)}(E)$.

Proof. Let us begin by making the observation that by construction E - as it is elementary - can be written as a finite collection of boxes, and further, given that E is elementary, it is necessarily bounded.
Let $\epsilon > 0$ be given. Let us consider two cases then, the first where E is closed and the second where it is not. Beginning with the first case, let $\{B_n\}_{n=1}^{\infty}$ be a cover of E satisfying $\sum_{n=1}^{\infty} \leq \lambda^*(E) + \frac{\epsilon}{2}$. To ensure that this cover is open, let us enlarge each B_n by $\frac{\epsilon}{2^{n+1}}$, and call this open cover $\{B'_n\}_{n=1}^{\infty}$. Thus we have that

$$\sum_{n=1}^{\infty} |B'_n| \leq \sum_{n=1}^{\infty} |B_n| + \sum_{n=1}^{\infty} \frac{\epsilon}{2^{n+1}} \leq \lambda^*(E) + \epsilon$$

Now observe that since E is closed and bounded, it is compact by Heine-Borel. As a consequence of this result, and the fact that $\{B'_n\}_{n=1}^{\infty}$ is an open cover, there exists N such that $E \subseteq \cup_{n=1}^{N} B'_n$. Applying the elementary measure operator to each side, we now have that

$$m^{(E)}(E) \leq m^{(E)}(\cup_{n=1}^{N} B'_n) = \sum_{n=1}^{N} |B'_n| \leq \sum_{n=1}^{\infty} |B'_n|$$

From this and the above, we now have that $m^{(E)}(E) \leq \lambda^*(E) + \epsilon$. Since $\lambda^*(E)$ is obviously bounded from below by elementary measure, the result is shown.
Now consider the other case, where it is not closed. By definition of an elementary set, E can be written as $\cup_{n=1}^{k} B_n$. In each of these, we can cull each of the boxes by $\frac{\epsilon}{k}$ to ensure that they are closed. Calling these new sets B'_n we make use of the prior case to show that $\lambda^*(\cup_{n=1}^{k} B'_n) = m^{(E)}(\cup_{n=1}^{k} B'_n) = m^{(E)}(B'_1) + \cdots + m^{(E)}(B'_k) \geq m^{(E)}(B_1) + \cdots + m^{(E)}(B_k) - \epsilon = m^{(E)}(E) - \epsilon$. Since ϵ was arbitrary, the result is shown.
Hence we have that $\lambda^*(E) = m^{(E)}(E)$. ■

An immediate and important consequence of this result is that $m_{(J,*)}(E) \leq \lambda^*(E) \leq m^{(J,*)}(E)$.[12]

The natural next question how do we go from a set's Lebesgue *outer* measure to a set's Lebesgue measure? To do this requires that we set a criterion for Lebesgue measurability. There are a number of possible criterions, but here are perhaps the two most common:

1. A set $E \subseteq \mathbb{R}^n$ is Lebesgue measurable if $\forall \epsilon > 0$, there is an open set $U \subseteq \mathbb{R}^n$ with $E \subseteq U$ that satisfies $\lambda^*(U \setminus E) \leq \epsilon$, and further we say that E has Lebesgue measure $\lambda(E) = \lambda^*(E)$.

2. (Caratheodory's criterion) A set $E \subseteq \mathbb{R}^n$ is Lebesgue measurable if and only if $\forall V \subseteq \mathbb{R}^n$ $\lambda^*(V) = \lambda^*(V \cap E) + \lambda^*(V \cap E^c)$

The intuition behind the first is simply that a set is Lebesgue measurable if we can surround it in open sets to an arbitrary level of precision. The second definition is almost completely unintuitive, but is far more important in general measure theory, because it abstracts more easily than the first one.

In either case, we can immediately see that certain classes of sets are $\lambda-$measurable.

[12] To see this, simply look at the definition of both Jordan Inner measure and Jordan Outer measure.

(i) Every open set

(ii) Every closed set

(iii) Every set E such that $\lambda^*(E) = 0$

(iv) $\emptyset$ is measurable

(v) If $E \subseteq \mathbb{R}^n$ is $\lambda-$measurable, then so is $\mathbb{R}^n \setminus E$.

(vi) Let $\{E_n\}_{n=1}^{\infty}$ be a collection of $\lambda-$measurable sets, then $\cup_{n=1}^{\infty} E_n$ is $\lambda-$measurable.

(vii) Let $\{E_n\}_{n=1}^{\infty}$ be a collection of $\lambda-$measurable sets, then $\cap_{n=1}^{\infty} E_n$ is $\lambda-$measurable.

One should note that conditions $(iv), (v)$, and (vi) imply that the collection of $\lambda-$measurable sets forms a $\sigma-$algebra.

To provide a precursor of sorts to some more important results in the theory of measurable functions later on, we are going to prove the *Dominated Convergence Theorem for Measurable Sets*. To prove this, however, we need to first prove two monotone convergence results.

Upward Monotone Convergence for $\lambda-$measurable sets: *If $E_1 \subseteq E_2 \subseteq \cdots \subset \mathbb{R}^n$ is a countable sequence of $\lambda-$measurable sets, then $\lambda(\cup_{n=1}^{\infty} E_n) = \lim_{n\to\infty} \lambda(E_n)$.*

Proof. If there exists N such that $\lambda(E_N) = \infty$, then the result is trivial. So without loss, let us consider the case where $\lambda(E_n)$ is finite for all $n \in \mathbb{N}$. Begin by observing that we can write $\cup_{n=1}^{\infty} E_n$ as the countable union of lacunae. That is to say that we can write $\cup_{n=1}^{\infty} E_n$ as $\cup_{n=1}^{\infty}(E_n \setminus \cup_{k=1}^{n-1} E_k)$. Applying λ to both sides, we now have that

$$\begin{aligned}
\lambda(\cup_{n=1}^{\infty} E_n) &= \lambda(\cup_{n=1}^{\infty}(E_n \setminus \cup_{k=1}^{n-1} E_k)) \\
&= \sum_{n=1}^{\infty} \lambda(E_n \setminus E_{n-1}) \\
&= \sum_{n=1}^{\infty} \lambda(E_n) - \lambda(E_{n-1}) \\
&= \lim_{n\to\infty} \lambda(E_n)
\end{aligned}$$

Where the first step follows from the upwards containment hypothesis and the third from the fact that this is a telescoping series. ■

Downward Monotone Convergence for* $\lambda-$*measurable sets: *If $\mathbb{R}^n \supset E_1 \supseteq E_2 \supseteq \dots$ is a countable sequence of $\lambda-$measurable sets, with at least one having finite Lebesgue measure, then $\lambda(\cap_{n=1}^{\infty} E_n) = \lim_{n\to\infty} \lambda(E_n)$.*

Proof. First observe that if at least one of the sets has finite measure, then the measure of the intersection is necessarily finite as well.[a] By containment and the fact that we are considering intersection then, we can truncate our sequence of sets to only those with finite measure. Without loss, let us consider this truncation.
Observe that $\lambda(E_1 \setminus \cap_{n=1}^{\infty} E_n) = \lambda(E_1) - \lambda(\cap_{n=1}^{\infty} E_n)$. Now observing that $E_1 \setminus \cap_{n=1}^{\infty} E_n = \cup_{n=1}^{\infty} E_n \setminus E_{n+1}$, we have that $\lambda(E_1) - \lambda(\cap_{n=1}^{\infty} E_n) = \lambda(\cup_{n=1}^{\infty} E_n \setminus E_{n+1}) = \sum_{n=1}^{\infty} \lambda(E_n) - \lambda(E_{n+1})$. As this is again a telescoping series, we have that $\sum_{n=1}^{\infty} \lambda(E_n) - \lambda(E_{n+1}) = \lambda(E_1) - \lim_{n\to\infty} \lambda(E_{n+1})$. Canceling $\lambda(E_1)$, we have that

$$\lambda(\cap_{n=1}^{\infty} E_n) = \lim_{n\to\infty} \lambda(E_{n+1})$$

Which was what was wanted. ■

[a] If one is not finite, then the above claim is not true. Consider the sequence of sets formed by removing from $\mathbb{R}^n$ $n-$balls with radius equal to the index.

With both of these results in hand, we can now prove the desired result. Recall that we say that a sequence of sets $\{E_n\}$ converges to a set E if the indicator functions 1_{E_n} converges pointwise to 1_E. Likewise a sequence of sets converges if and only if its limit superior and limit inferior coincide and they are equal to E.[13]

[13]

$$\limsup_{n\to\infty} E_n = \cap_{n=1}^{\infty}(\cup_{k=n}^{\infty} E_k)$$

$$\liminf_{n\to\infty} E_n = \cup_{n=1}^{\infty}(\cap_{k=n}^{\infty} E_k)$$

Dominated Convergence Theorem for $\lambda-$measurable sets: *If $\{E_n\}_{n=1}^{\infty}$ be a sequence of $\lambda-$measurable sets such that $E_n \subseteq F$ $\forall n$, with F having finite Lebesgue measure, and $\{E_n\} \to E$, then $\lambda(E_n) \to \lambda(E)$.*

Proof. Begin by supposing that $\{E_n\}_{n=1}^{\infty}$ is a sequence of $\lambda-$measurable sets such that $E_n \subseteq F$ $\forall n$, with F having finite Lebesgue measure, and $\{E_n\} \to E$. As was stated above, the convergence of $\{E_n\}$ to E gets us that the limit superior and limit inferior coincide and further that they are equal to E.
Observing that $\cap_{n=1}^{\infty}(\cup_{k=n}^{\infty} E_k)$ is a decreasing sequence of sets, and that the measure of this set is bounded by $\lambda(F)$, given the condition that $E_n \subseteq F$ $\forall n$, we have by the *Downward Monotone Convergence for λ-Measurable Sets* result, that $\lambda(\cap_{n=1}^{\infty}(\cup_{k=n}^{\infty} E_k)) = \lim_{n\to\infty} \lambda(\cup_{k=n}^{\infty} E_k))$. Now, since $\cup_{k=n}^{\infty} E_k \supseteq E_n$, we have that $\lambda(\cup_{k=n}^{\infty} E_k) \geq \lambda(E_n)$, which together means

$$\lambda(E) = \lambda(\cap_{n=1}^{\infty}(\cup_{k=n}^{\infty} E_k)) = \lim_{n\to\infty} \lambda(\cup_{k=n}^{\infty} E_k) \geq \lim_{n\to\infty} \lambda(E_n)$$

.
Mirroring this, we observe that $\cup_{n=1}^{\infty}(\cap_{k=n}^{\infty} E_k)$ constitutes an increasing sequence of sets, and thus by the *Upward Monotone Convergence for λ-Measurable Sets* result, we have that

$$\lambda(E) = \lambda(\cup_{n=1}^{\infty}(\cap_{k=n}^{\infty} E_k)) = \lim_{n\to\infty} \lambda(\cap_{k=n}^{\infty} E_k) \leq \lim_{n\to\infty} \lambda(E_n)$$

where the last step follows from the fact $\cap_{k=n}^{\infty} E_k \subseteq E_n$ and the monotonicity of Lebesgue measure. Taken together, these imply that

$$\lim_{n\to\infty} \lambda(E_n) = \lambda(E)$$

■

In addition to the typical properties that we want from a measure - namely translation invariance, non-negativity, and countable additivity for disjoint sets - the Lebesgue measure has several additional useful properties, including that it is **Inner Regular**[14], which is the condition that:

$$\mu(E) = \sup_{K \subseteq E, K \text{ compact}} \mu(K)$$

This means that we can approximate the measure of a set by considering some compact subset. To do this, we need first prove a lemma about Lebesgue measurability.

[14]Indeed, Lebesgue measure is also outer regular, in the sense that it is by construction an infenum of coverings by open sets, and further it is locally finite in the sense that every point in the underlying space has a neighborhood around it with finite measure. Inner regularity, outer regularity, and local finiteness together define what is called a Radon measure, if your space of interest is Hausdorff, and your $\sigma-$algebra is Borel.

Criterion for Lebesgue Measurability: *A set $E \subseteq \mathbb{R}^n$ is Lebesgue Measurable if and only if $\forall \epsilon > 0$, there exists a closed set K contained in E satisfying $\lambda^*(E \setminus K) < \epsilon$.*

Proof. Let $E \subseteq \mathbb{R}^n$ be Lebesgue measurable. From this, we know that E^c is also Lebesgue measurable, which implies that $\forall \epsilon > 0$ we can generate an open cover, call it U_c around E_c such that $\lambda^*(U_c \setminus E_c) < \epsilon$. Let K be the complement of U_c and observe that because U_c is an open cover, K is closed. Now observe that $U_c \setminus E_c = U_c \cap E = E \setminus K$.[a] Hence we have that $\lambda^*(E \setminus K) = \lambda^*(U_c \setminus E_c) < \epsilon$. ■

[a]Let $x \in U_c \setminus E_c$, then $x \in U_c$ and $x \notin E_c$. This implies that $x \notin (U_c)_c = K$ and $x \in E$, and thus that $x \in E \setminus K$.

With this result in hand, we can now move on to show that $\lambda(\cdot)$ is inner regular.

Claim: *$\lambda(\cdot)$ is an inner regular measure.*

Proof. Let E be a $\lambda-$measurable set. We have two cases of interest. The first is when E is bounded. This result follows trivially from the prior lemma.
The second case is when E is unbounded. Let $\epsilon > 0$ be given.
This is again going to have two cases, one where the measure of the set is infinite and one where it is finite[a] For both cases, however, we are going to need special sequence of sets. Namely, let A_i be given by $E \cap \bar{B}(0, i)$ - the intersection of a closed $n-$ball of radius i with the set E. As the intersection of $\lambda-$measurable sets is $\lambda-$measurable as well, each of the elements of this sequence is $\lambda-$measurable. By the upward monotone convergence result from before, we have that $\lambda(\cup_{i=1}^{\infty} A_i) = \lambda(E) = \lim_{i\to\infty} A_i = \infty$. This gets us the unbounded result for the case with infinite measure.
For the case with $\lambda(E) < \infty$, observe that from the monotone convergence result, we have that $\lambda(E) = \lim_{n\to\infty} A_i$. This implies that we can pick $i \in \mathbb{N}$ large enough such that $\lambda(E \setminus A_i) < \frac{\epsilon}{2}$. Observing that A_i is by construction $\lambda-$measurable and bounded, we can find another compact set $K \subseteq A_i$ such that $\lambda^*(A_i \setminus K) < \frac{\epsilon}{2}$. Taken together then, we have that $\lambda(E \setminus K) < \epsilon$. ■

[a]To understand how this latter subcase might occur, consider, say $\mathbb{N} \cup [0, 1]$. This set is obviously unbounded. This example also provides a bit of a clue as to how we want to prove the result, once you realize that what you can do is effectively just chop off countably many points and retain a set of equivalent measure.

2.3 A Brief Digression on Modes of Convergence

In order to talk about Lebesgue measurable functions, we need some way to talk about a sequence of functions converging to a function. There are a number of possibilities. For the sake of brevity here, let us fix a sequence of functions $\{f_i\}$, such that for all $i \in \mathbb{N}$ $f_i : \mathbb{R}^n \to \mathbb{R}$. That is to say that we have a sequence of functions with common domain.

You will see in the next section that we are initially concerned only with two. We say $\{f_i\}$ is pointwise convergent to f (or **converges pointwise** to f), if

$$\forall \epsilon > 0\ \forall x \in X\ \exists N_{\epsilon,x} \text{ such that} |f_i(x) - f(x)| < \epsilon\ \forall i \geq N_{\epsilon,x}$$

In contrast to this, we say that $\{f_i\}$ is uniformly convergent to f (or **converges uniformly** to f) if

$$\forall \epsilon > 0 \; \exists N_\epsilon \text{ such that} |f_i(x) - f(x)| < \epsilon \; \forall x \in X \; \forall i \geq N_\epsilon$$

These two modes of convergence are similar, but very much not the same. The distinction is roughly the same distinction between continuity and uniform continuity. Recall when we are concerned with continuity of a function, we are concerned with continuity at a point. Thus, when given ϵ, we can choose N based on the point we are at. In contrast, when we are considering uniform continuity, we are concerned with the entire domain of the function, and thus our choice of ϵ cannot depend on a specific part of the function.

Pointwise convergence is like continuity in the strict sense that we can allow our N to vary depending on the x we are considering the behavior of the sequence of functions at. Uniform convergence is like uniform continuity in the sense that we cannot do that.

To make another thing explicit, as uniform continuity implies continuity, so too does uniform convergence imply pointwise convergence. Similarly, there are continuous functions that are not uniformly continuous, and there are sequences of functions that converge pointwise but not uniformly.

One natural generalization of these two modes of convergence is to think about the behavior of the sequence of functions on sets of different sizes. For instance, one is often interested in claims that are made about functions almost everywhere: we say that a property $P(x)$ holds **almost everywhere** if $\mu(x \in X | \text{P(x) is not true}) = 0$.[15] In words, we say something is true almost everywhere if the set where it is not true has measure zero.[16]

2.4 Lebesgue Measurable Functions and the Lebesgue Integral

With our results about Lebesgue measurability for sets in hand, we now turn to Lebesgue measurable *functions*.

To understand this, we want to briefly introduce the concept of simple functions. The easy comparison here is that a simple function is to a function what an elementary set is to a set. So, as an elementary set is simply a finite collection of boxes, an unsigned simple function is simply a function $f : \mathbb{R}^n \to [0, \infty]$ such that

$$f = \sum_{i=1}^{m} c_i 1_{E_i} \; \textit{with } E_i \textit{ Lebesgue measurable} \; \forall i$$

In other words, an unsigned simple function is one that can be written as a finite combination of scaled indicator functions on λ-measurable sets.[17]

One reason to care about simple functions precisely because we define the class of **Lebesgue Measurable** functions as those unsigned functions $f : \mathbb{R}^n \to [0, \infty]$ that are the pointwise limit of unsigned simple functions. There are, however, a number of other equivalent definitions with

[15] $\mu(\cdot)$ here is taken by an arbitrary measure.

[16] This turns out to be very useful. Specifically, we can use almost everywhere equality to define a set of equivalence classes which partition the space of functions defined on some domain.

[17] Recall that an indicator function is simply

$$\not\Vdash_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}$$

perhaps the most significant one being analogous to continuous functions. To see what I mean, recall the topological definition of continuity. Let $f : \mathbb{R}^n \to \mathbb{R}$ be given: we say that f is continuous if $\forall U \subseteq \mathbb{R}$, with U open, $f^{-1}(U) = V$, for some open set $V \subseteq \mathbb{R}^\ltimes$. That is to say that a function is continuous if and only if the preimage of an open set under f is open. Measurability, like continuity is about preserving some structure from the image of the function. But where continuity is concerned with open sets, measurability is concerned with measurable sets.

With this in mind, we say that a function $f : \mathbb{R}^n \to [0, \infty]$[18] is **Lebesgue Measurable** if it satisfies any of the following equivalent criteria[19]:

1. f is Lebesgue measurable.
2. f is the pointwise limit of unsigned simple functions f_n.
3. f is the pointwise almost everywhere limit of unsigned simple functions f_n.
4. f is the supremum $f(x) = \sup_n f_n(x)$ of an increasing sequence $0 \leq f_1 \leq f_2 \leq \ldots$ of unsigned simple functions f_n, each of which are bounded with finite measure support.[20]
5. $\forall \tau \in [0, \infty], \{x \in \mathbb{R}^n : f(x) > \tau\}$ is Lebesgue measurable.
6. $\forall \tau \in [0, \infty], \{x \in \mathbb{R}^n : f(x) \geq \tau\}$ is Lebesgue measurable.
7. $\forall \tau \in [0, \infty], \{x \in \mathbb{R}^n : f(x) < \tau\}$ is Lebesgue measurable.
8. $\forall \tau \in [0, \infty], \{x \in \mathbb{R}^n : f(x) \leq \tau\}$ is Lebesgue measurable.
9. Let I be an interval in $[0, \infty)$, then f is Lebesgue measurable if $f^{-1}(I)$ is Lebesgue measurable.
10. Let U be a relatively open set in $[0, \infty)$, then f is Lebesgue measurable if $f^{-1}(U)$ is Lebesgue measurable.
11. Let K be a relatively closed set in $[0, \infty)$, then f is Lebesgue measurable if $f^{-1}(K)$ is Lebesgue measurable.

Indeed, one claim that one should see is true almost immediately is that all continuous functions $f : \mathbb{R}^n \to \mathbb{R}_+$ are Lebesgue measurable.

[18]Recall again that this set is the positive *extended* reals.

[19]The caveat to these definitions is that they are somewhat idiosyncratic to $\mathbb{R}^n$. In particular, any definition making use of the words open or closed is going to generalize in a bit of different manner. The reason that they work in this context is because it turns out that the Lebesgue $\sigma-$algebra is the completion of the smaller Borel $\sigma-$algebra, which is the minimal $\sigma-$algebra generated by the collection of all open intervals. If that doesn't make sense right now, that's ok, it will be developed in a bit more detail later when we move on from measure theory in $\mathbb{R}^n$ and into abstract measure spaces. As another aside, this little idiosyncrasy in the definition of the Lebesgue measurability introduces some complications where statements we want to make in general are not true. For example, it is not true that the composition of Lebesgue measurable functions need be Lebesgue measurable. I leave this as a (difficult) exercise for the reader.

[20]Recall that the support of a function $f : X \to \mathbb{R}$ is given by $\{x \in X | f(x) \neq 0\}$

Claim: *All continuous functions from $\mathbb{R}^n$ to $\mathbb{R}_+$ are measurable.*

Proof. Let U be a relatively open set in $[0, \infty)$. Since f is continuous, the preimage of U under f must be an open set as well. Since all the open sets are Lebesgue measurable, the result is shown. ■

We might also be interested in the graph of a function.

Claim: *Let $f : \mathbb{R}^n \to [0, \infty)$ be an unsigned measurable function. Then the graph of f, $\Gamma(f, \mathbb{R}^n) \equiv \{(x, f(x)) | \, x \in \mathbb{R}^n\}$ has Lebesgue measure zero.*

Proof. First we shall make an observation that $\mathbb{R}^n$ can be written as a countable union of almost disjoint boxes, so if we can prove this on a box, then we can use this fact to expand it to all of $\mathbb{R}^n$.[a] Keeping this mind, let us now consider f on some box $E \subseteq \mathbb{R}^n$, with E having finite measure.
Let $\epsilon > 0$ be given. Define E_m by $\{x \in E | \epsilon m \leq f(x) < \epsilon(m+1)\}$ and let $m = 0, 1, 2, \ldots$.[b] Observe that since f is unsigned, $\cup_{m=0}^{\infty} E_m = E$. Now observe that $\lambda^*(\Gamma(f, E_m)) \leq \epsilon \lambda^*(E_m)$. From countable subadditivity and this observation, we have that

$$\lambda^*(\Gamma(f, \cup_{m=0}^{\infty} E_m) \leq \sum_{m=0}^{\infty} \lambda^*(\Gamma(f, E_m)) \leq \epsilon \sum_{m=0}^{\infty} \lambda^*(E_m) = \epsilon \lambda^*(E)$$

As ϵ was arbitrary and $\lambda^*(E) < \infty$ by assumption the claim follows on E. And since $\mathbb{R}^n$ can be written as a countable union of such boxes, the result follows on all of $\mathbb{R}^n$.
Since all sets with Lebesgue outer measure zero are Lebesgue measurable, it follows that the graph of a measurable function is a null set. ■

[a] The fact that $\mathbb{R}^n$ can be written as a countable union of sets of finite measure is a very important property called σ-finiteness.

[b] Observe that since f is Lebesgue measurable, E_m is $\lambda-$measurable for all m, as it is the intersection of sets in criteria 6 and 7 above, and the intersection of $\lambda-$measurable sets is itself $\lambda-$measurable.

Claim: *Let $f : \mathbb{R}^n \to [0, \infty)$ be an unsigned measurable function. Then the set $\{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t \leq f(x)\}$ is λ-measurable.*

Proof. We begin the proof by observing that if f is an unsigned simple function, the result is obviously true. Simply consider the (finite) number of values that f takes on, given that it is a simple function, and partition $\mathbb{R}^n$ based on the indicator functions that make up the unsigned simple function. Observing that this function is everywhere equivalent to a simple function g with m components, one for each of the finite number of values in the range of g, we can without loss take f to be finite summation of m disjoint indicator functions.[a] Then $\{(x,t)\}$ can be written as the disjoint union of λ-measurable sets, as the cartesian product of λ-measurable sets is λ-measurable.[b]
Now let f be an unsigned measurable function with domain in $\mathbb{R}^n$. Observe that since f is an unsigned Lebesgue measurable function, it is the pointwise limit of unsigned simple functions f_i, and further $f_i \leq f_j \; \forall i < j$.
It remains to show that

$$\cup_{i=1}^{\infty}\{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t < f_i(x)\} = \{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t < f(x)\}$$

Let $(x_0, t_0) \in \cup_{i=1}^{\infty}\{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t < f_i(x)\}$. Then there exists $m \in \mathbb{N}$ such that $t_0 < f_m(x_0)$. Since $f_i(x_0) \leq f(x_0) \; \forall i \in \mathbb{N}$, we have then that $t_0 < f(x_0)$, so $(x_0, t_0) \in \{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t < f(x)\}$.
Now let $(x_1, t_1) \in \{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t < f(x)\}$. Suppose that $(x_1, t_1) \notin \cup_{i=1}^{\infty}\{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t < f_i(x)\}$. Then $t_1 \geq f_m(x_1) \; \forall m \in \mathbb{N}$, so t_1 is by definition an upper bound on $\{f_i(x_1)\}$. But $f(x_1) = \sup_i f_i(x_1)$, as f is Lebesgue measurable. Since $t_1 < f(x_1)$ we have a contradiction, so $(x_1, t_1) \in \cup_{i=1}^{\infty}\{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t < f_i(x)\}$.
By double containment, we have that $\cup_{i=1}^{\infty}\{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t < f_i(x)\} = \{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t < f(x)\}$.
Since the countable union of Lebesgue measurable sets is Lebesgue measurable, we have that $\{(x,t)|x \in \mathbb{R}^n \wedge 0 \leq t < f(x)\}$ is Lebesgue measurable. Since by the prior result we know that the graph of a measurable function is a null set, and the union of Lebesgue measurable sets is itself Lebesgue measurable, the claim is shown. ∎

[a] For clarity, I say that $\not\Vdash_{E_i}$ is disjoint with $\not\Vdash_{E_j}$ if $E_i \cap E_j = \emptyset$.

[b] Note that I am abusing notation here. Technically, we are dealing with 3 different Lebesgue measures here: one on $\mathbb{R}^n$, one on $\mathbb{R}$, and one on $\mathbb{R}^{n+1}$. When I am saying a set is λ-measurable, I mean the set is measurable with respect to the appropriate Lebesgue measure on the space the set lives in.

From the simple functions, we can define the **Simple Integral** by

$$Simp \int_{\mathbb{R}^n} f(x)dx \equiv \sum_{i=1}^{m} c_i \lambda(E_i)$$

Now let f be an unsigned but necessarily measurable function from $\mathbb{R}^n$ to $[0, \infty]$. Using the simple integral, we can define the **Lower Unsigned Lebesgue Integral** by

$$\underline{\int_{\mathbb{R}^n}} f(x)dx \equiv \sup_{0 \leq g \leq f \text{ g simple}} Simp \int_{\mathbb{R}^n} g(x)dx$$

Likewise, we can define the **Upper Unsigned Lebesgue Integral** by

$$\overline{\int_{\mathbb{R}^n}} f(x)dx \equiv \inf_{f \leq g \text{ g simple}} Simp \int_{\mathbb{R}^n} g(x)dx$$

There are a few properties of the lower Lebesgue integral that are useful.

Claim: *The Lower Lebesgue Integral is homogeneous of degree 1.*

Proof. To show that the lower integral is homogeneous of degree 1, we need simply show that

$$\forall \alpha \in \mathbb{R}_+ \underline{\int_{\mathbb{R}^n}} \alpha f(x)dx = \alpha \underline{\int_{\mathbb{R}^n}} f(x)dx$$

Observe that

$$\underline{\int_{\mathbb{R}^n}} \alpha f(x)dx = \sup_{0 \leq g \leq \alpha f} Simp \int_{\mathbb{R}^n} g(x)dx = \alpha \left[\sup_{0 \leq h \leq f} Simp \int_{\mathbb{R}^n} h(x)dx \right] = \alpha \underline{\int_{\mathbb{R}^n}} f(x)dx$$

Where the first and third equalities follow from the definition of the Lower Lebesgue Integral, and the second equality follows from the homogeneity of the supremum operator. ■

The fact that we are building up the Lebesgue integral out of simple functions should reveal that it is in effect very different from the Riemann integral. Recall the construction of the Riemann Integral requires one to first construct a mesh on the domain. Within each cell of the mesh, we then take the infenum and supremum of the function to be integrated, and then we calculate the upper and lower Darboux sums. The Riemann Integral is, to a first approximation, the result of taking the size of cells in the mesh to 0. By building the Lebesgue integral out of simple functions, we are essentially partitioning the *range* of the function.

In any event, we are now in position to define the **Lebesgue Integral**. Specifically, let f be a measurable function from $\mathbb{R}^n$ to $[0, \infty]$. Then we say that the (unsigned) Lebesgue integral is given by the lower Lebesgue integral:

$$\int_{\mathbb{R}^n} f(x)dx = \underline{\int_{\mathbb{R}^n}} f(x)dx$$

There are a couple of very nice results that I will state here that are in general useful.

Claim: *A function $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ is Riemann Integral on $[a, b]$ if and only if its set of discontinuities has (Lebesgue) measure zero.*

Claim: *Let $f : [a,b] \to [0,\infty]$ be Riemann Integrable. Extending f to cover all of $\mathbb{R}$ by setting $f(x) = 0 \forall x \notin [a,b]$, then the Riemann integral of f coincides with the Lebesgue Integral $\int_{\mathbb{R}^n} f(x)dx$.*[a]

Proof. This "proof" is not a proof, it's more of a discussion of the idea, since this is not relevant to what you need first year. One way to approach it is to use the equivalence of Riemann and Darboux integrability and observe that all step functions are simple functions and vice versa. ■

[a]This claim is not true for improper Riemann integrals. The canonical counterexample is $\frac{sin(x)}{x}$ on $\mathbb{R}$.

The latter result is nice because we have a well developed toolkit for calculating the values of Riemann integrals, so we can simply use those, at least for *proper* Riemann integrals.

We will also now get a fairly important result.

Markov's Inequality: *Let $f : \mathbb{R}^n \to [0,\infty]$ be a Lebesgue-measurable function. Then*

$$\forall \tau \in (0,\infty) \ \lambda(\{x \in \mathbb{R}^n : f(x) \geq \tau\}) \leq \frac{1}{\tau} \int_{\mathbb{R}^n} f(x)dx$$

Proof. Observe that $\forall x \in \mathbb{R}^n \ \tau 1_{\{x \in \mathbb{R}^n | f(x) \geq \tau\}} \leq f(x)$.[a] Integrating both sides, we recover that

$$\int_{\mathbb{R}^n} \tau 1_{\{x \in \mathbb{R}^n | f(x) \geq \tau\}} dx \leq \int_{\mathbb{R}^n} f(x)dx$$

From the homogeneity of the Lebesgue integral, we then have that

$$\tau \int_{\mathbb{R}^n} 1_{\{x \in \mathbb{R}^n | f(x) \geq \tau\}} dx \leq \int_{\mathbb{R}^n} f(x)dx$$

Dividing both sides by τ[b] and observing that the Lebesgue integral over a characteristic function is exactly the measure of the set, we now have that

$$\lambda(\{x \in \mathbb{R}^n | f(x) \geq \tau\}) \leq \frac{1}{\tau} \int_{\mathbb{R}^n} f(x)dx$$

■

[a]To see why this is true, consider the case where the indicator function takes on the value 1 and when it does 0. If the indicator function is equal to 1, then by construction it must be that $\tau \leq f(x)$. If the indicator function is 0, then because $\tau * 0 = 0$ and the range of $f(x)$ is $[0,\infty]$, we have the desired inequality.

[b]Which is admissible because by assumption $\tau \in \mathbb{R}_{++}$

For our purposes, we are going to be concerned with a subset of the measurable functions, specifically those f which are absolutely integrable.

Specifically, we say that an unsigned $\lambda-$measurable function f is **absolutely integrable** if it satisfies

$$||f||_{L_1(\mathbb{R}^n)} \equiv \int_{\mathbb{R}^n} |f(x)|dx < \infty$$

This quantity $||f||_{L_1(\mathbb{R}^n)}$ is called the $L_1(\mathbb{R}^n)$ **norm**[21] **of** f.[22]

In general, the $L_p(\mathbb{R}^n)$ **norm of** f is given by

$$||f||_{L_p(\mathbb{R}^n)} \equiv \left(\int_{\mathbb{R}^n} |f(x)|^p dx\right)^{\frac{1}{p}}$$

.

There is a much deeper connection between absolutely integrable functions and continuous functions, given by the following (famous) result.

Lusin's Theorem: *Let $f : \mathbb{R}^n \to \mathbb{C}$ be absolutely integrable. Let $\epsilon > 0$ be given. Then there exists a $\lambda-$measurable set $E \subset \mathbb{R}^m$ of measure at most ϵ, such that the restriction of f to the complementary set $\mathbb{R}^n \setminus E$ is continuous on that set.*

This theorem does not say that the function is continuous off the null set. What it says is that if we remove the null set from the domain of the function, then we have a function that is everywhere continuous.[23]

In combination with the inner regularity of the Lebesgue measure, and another strong result called Tietze's Extension theorem, that you will perhaps see in Mini 3 of Micro, we have the result that if you have a measurable function f, you can make it continuous simply by changing the values on the null set.

I leave as an exercise demonstrating that the L_p norm is a (semi)-norm.[24]

Now we will move on to more abstract spaces.

2.5 Measure Theory in Abstract Measure Spaces

We will begin by reciting some definitions:

Let X be an arbitrary set.

We say that a collection of subsets $\mathcal{F}$ is a **Boolean Algebra**[25] if it satisfies the following:

1. $\emptyset \in \mathcal{F}$
2. $E \in \mathcal{F} \implies X \setminus E \in \mathcal{F}$
3. $\{E_i\}_{i=1}^n \implies \cup_{i=1}^n E_i \in \mathcal{F}\ \forall n \in \mathbb{N}$

It turns out that these simple algebras are not in general strong enough for us, so we want to strengthen the above concept to what is called a $\sigma-$**algebra** on X, which is a collection of subsets Σ satisfying the following:

1. $\emptyset \in \Sigma$

[21] A norm is simply a generalization of the length of a vector. Norms induce metrics, and metrics induce topologies. Similarly, semi-norms induce pseudo-metrics. In the case of L_p norms, we can make the (semi)-norm into a norm by changing our notion of equality from pointwise equality to almost everywhere equality.

[22] It is common to omit the dimensional indication when referring to the norm, as it is generally obvious the domain of the functions under consideration.

[23] Consider, say, the Dirichlet function. If you restrict your consideration to $\mathbb{R} \setminus \mathbb{Q}$, then you recover a function that is identically 0 at every irrational. The Dirichlet function itself, however, is continuous nowhere.

[24] A semi-norm is a norm which is not separating. That is to say that $\exists f$ such that $||f||_{L_p} = 0$, but $f \neq 0$, where 0 here refers to the 0 function.

[25] Or sometimes we simply call it an **Algebra**.

2. $A \in \Sigma \implies X \setminus A \in \Sigma$
3. $A_1, A_2, \dots \in \Sigma \implies \cup_{i=1}^{\infty} A_i \in \Sigma$

In other words, we want to strengthen the finite union requirement of the Boolean Algebra to the countable union one.

In either case, one should immediately notice that DeMorgan's laws[26] implies that any $\sigma-$algebra will also be closed under countable intersection.

As this definition is somewhat opaque, let us consider two examples: one of a set of subsets that is a $\sigma-$Algebra and one that is not.

So let $X = \{1, 2, 3, 4\}$.

Let us takes as candidates the following:

$$\mathcal{F} = \{\emptyset, \{1, 3\}, \{2\}, \{1, 2, 3, 4\}\}$$

$$\mathcal{G} = \{\emptyset, \{1\}, \{2, 3, 4\}, \{1, 2, 3, 4\}\}$$

Now let us observe that $\mathcal{F}$ is clearly not a $\sigma-$algebra. Why? Well it fails criterion 2 because $\{2, 4\} \notin \mathcal{F}$. It also fails criterion 3 because $\{1, 2, 3\} \notin \mathcal{F}$.

I leave as a (trivial) exercise to the reader proving that $\mathcal{G}$ satisfies conditions $1 - 3$.

To some extent, there is an equivalence between a σ-algebra and a partition of the space. One way to see this is to think of the two most different σ-algebras on some space X. Observe that the smallest σ-algebra that we can generate is in fact $\{\emptyset, X\}$. This is what is called the **trivial σ-algebra**. On the other hand, we could break the set apart into every individual element. In this case, it is easy to see that this generates a σ-algebra 2^X. Note the symmetry between this and the power set on X. That is because these two objects are in fact one and the same. We call this maximal σ-algebra the **discrete σ-algebra**. The difference between these two algebras also serves to demonstrate one way of categorizing the size of σ-algebra. Specifically, let Σ_1 and Σ_2 be two σ-algebras on some space X. We say that Σ_1 is finer than Σ_2 if $\forall E \in \Sigma_1$ we have that $E \in \Sigma_2$. On the other hand, if Σ_1 contains fewer sets than Σ_2, we say that Σ_1 is finer.

But now let us suppose that we are already working with a collection of sets $\mathcal{F}$ on some space X. We might be interested in knowing what is the smallest σ-algebra that we could possibly work with that still includes every element of $\mathcal{F}$. Such a σ-algebra we call the **σ-algebra generated by $\mathcal{F}$**, and we denote it $< \mathcal{F} >$.

Claim: *Suppose that $\{\Sigma_i\}_{i=1}^{\infty}$ is a countable collection of (possibly repeating) σ-algebras. Then $\bigcap_{i=1}^{\infty} \Sigma_i$ is a σ-algebra.*

Proof. Let $\{\Sigma_i\}_{i=1}^{\infty}$ be a collection of σ-algebras. By definition, we have that X and $\emptyset$ are in each of the Σ_i, and are thus in their intersection. Now suppose that $E \in \bigcap_{i=1}^{\infty} \Sigma_i$. Then $\forall i \; E \in \Sigma_i$, and since σ-algebras are closed under complements, so too is $X \setminus E$. Hence $X \setminus E \in \bigcap_{i=1}^{\infty} \Sigma_i$. Now let us suppose that $\{E_j\}$ is a countable sequence of sets in $\bigcap_{i=1}^{\infty} \Sigma_i$. Then it is also a countable sequence of sets in each of E_j, so exploiting their own closure under countable union properties, we get that $\bigcup_{j=1}^{\infty} E_j \in \bigcap_{i=1}^{\infty} \Sigma_i$. Hence we have that the intersection of σ-algebras is itself a σ-algebra. ■

[26] $(A \cup B)^c = A^c \cap B^c \wedge (A \cap B)^c = A^c \cup B^c$

Note that this result says the intersection of σ-algebras is itself a σ-algebra. It is not in general true that the union of σ-algebras is a σ-algebra.

Formally, we define the σ-algebra generated by $\mathcal{F}$, which we denote $< \mathcal{F} >$, as the intersection of all σ-algebras that contain $\mathcal{F}$. Generated σ-algebras turn out to be very important. In particular, it is very common to find yourself concerned with the **Borel σ-Algebra**, which we typically notate with $\mathcal{B}$. This is the smallest $\sigma-$algebra that contains all the open sets in $\mathbb{R}^n$. Put differently, this is the σ-algebra generated by the collection of open subsets, where open subsets here are in reference to the usual topology[27] on $\mathbb{R}^n$.[28] The second is the **Lebesgue $\sigma-$algebra**, which we typically notate with $\mathcal{L}$. This is the $\sigma-$algebra consisting of the Lebesgue measurable sets. This is exactly the collection of sets that we considered in the first part of these notes. These two σ-algebras are very closely related, for reasons that will be discussed in slightly more detail later in the notes. Another special kind of σ-algebras is the restriction of a σ-algebra. Specifically, let $A \in \mathcal{X}$, then we define the **Restriction of $\mathcal{X}$ to** A, as

$$\mathcal{X} \upharpoonright_A \equiv \{A \cap E | E \in \mathcal{X}\}$$

Claim: *Let X be a space and let $\mathcal{X}$ be a σ-algebra on X. Let $A \in \mathcal{X}$. Then $\mathcal{X} \upharpoonright_A$ is a σ-algebra on A*

Proof. Observe that since $A \cap \emptyset = \emptyset$, we have that $\emptyset \in \mathcal{X}|A$. Likewise, suppose that $E \in \mathcal{X} \upharpoonright_A$. Then $E = U \cap A$ for some $U \in \mathcal{X}$. Since $\mathcal{X}$ is a σ-algebra, we have that $U^c \in \mathcal{X}$, and hence $U^c \cap A = E^c \in \mathcal{X} \upharpoonright_A$. Now suppose that $\{E_i\}$ is a countable collection of measurable sets in $\mathcal{X} \upharpoonright_A$. Then for each set, we can write E_i as $U_i \cap A$ for some $U_i \in \mathcal{X}$. Taking unions and exploiting the distributivity of intersection, we have that $\cup_{i=1}^{\infty}(U_i \cap A) = (\cup_{i=1}^{\infty} U_i) \cap A$. Since $\mathcal{X}$ is a σ-algebra and thus closed under countable union, it follows that $\cup_{i=1}^{\infty} E_i \in \mathcal{X} \upharpoonright_A$. ■

In general, we call the duple consisting of a set X and a $\sigma-$algebra $\mathcal{X}$ (i.e. $(X, \mathcal{X})$) a **measurable space.**

Now, let μ be a function from $\mathcal{X}$ to $\mathbb{R} \cup \{-\infty, \infty\}$ that satisfies the following properties:

1. $\forall B \in \mathcal{X}\ \mu(B) \geq 0$
2. $\mu(\emptyset) = 0$
3. $\mu(\cup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i)$ for all countable collections $\{B_i\}_{i=1}^{\infty}$, s.t. $B_i \cap B_j = \emptyset \forall i \neq j$

[27] For reference, a topology τ on a space X is a collection of sets that has the following properties:

1. $\emptyset, X \in \tau$
2. Let $\{T_i\}$ be an arbitrary collection of elements of τ. Then the union of the $\{T_i\}$ is itself an element of τ.
3. Let $\{T_i\}$ be a finite collection of elements of τ. Then the intersection of the $\{T_i\}$ is an element of τ.

We call the elements of a topology open sets.

[28] In general, given a space X and a topology on X (i.e. a topological space), the σ-algebra generated by the collection of all open sets according to that topology is called the Borel σ-algebra on X. In practice, every time we reference the Borel σ-algebra in this course, and probably anytime an economist references it, they are referring to the Borel σ-algebra on $\mathbb{R}^n$ with the usual topology. One should also see that the Borel σ-algebra is also generated by the collection of all *closed sets*, since the complement of an open set is a closed one and vice versa. Perhaps more famously, and probably most importantly for probability theory and the theory of distributions, the Borel σ-algebra on $\mathbb{R}$ is also generated by the collection of half open intervals of the form $(-\infty, a]\ \forall a \in \mathbb{R}$.

Then we say that μ is a **measure**, and those 3 criteria above are called the **Measure Axioms**.

In what is a somewhat obvious extension, if we impose the additional requirement that the measure of the whole space is unity - $\mu(X) = 1$ - we have a **probability measure.**

Further, we say that the triple $(X, \mathcal{X}, \mu)$ is a **measure space.**[29] There are a number of important properties that measure spaces might have, but two in particular are going to be relevant to this course. We say that a measure space $(X, \mathcal{X}, \mu)$ is finite if $\mu(X) < \infty$. It is I think obvious that all probability triples are going to be finite measure spaces. More generally, we say that a measure space $(X, \mathcal{X}, \mu)$ is σ-finite, if $\exists \{E_i\}, E_i \in \mathcal{X} \wedge \mu(E_i) < \infty \forall i$ and $X \subseteq \cup_{i=1}^{\infty} E_i$. In words, we say that a measure space is σ-finite if it can be covered entirely by a countable collection of sets, each with finite measure.

Claim: *Measure is monotonic in that sense that* $E \subseteq F \implies \mu(E) \leq \mu(F)$.

Proof. Let E and F be given with $E \subseteq F$. Define $G = F \setminus E$. Observing that $E \cup G = F$, we have that $\mu(E \cup G) = \mu(F)$. Since, by construction $G \cap E = \emptyset$, by finite disjoint additivity[a], we have that $\mu(E) + \mu(G) = \mu(F)$. By non-negativity of measure, we have that $\mu(G) \geq 0$, so it follows that $\mu(E) \leq \mu(F)$. ■

[a] This follows from the fact that we have countable disjoint additivity and that $\mu(\emptyset) = 0$

[29] Note that this is a **measure space**, not a **metric space**, which is a space equipped with some notion of distance.

Claim: *Let $(X, \mathcal{X}, \mu)$ be given. Then if $\{E_i\}$ is $\mathcal{X}-$measurable $\forall i \in \mathbb{N}$, we have*

$$\mu(\bigcup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} \mu(E_i)$$

Proof. Let $\{E_i\}$ be given as above. Observe that

$$\cup_{i=1}^{\infty} E_i = \cup_{i=1}^{\infty}(E_i \cap (\cup_{k=1}^{i-1} E_k)^c)$$

It follows then that

$$\mu(\cup_{i=1}^{\infty} E_i) = \mu(\cup_{i=1}^{\infty}(E_i \cap (\cup_{k=1}^{i-1} E_k)^c))$$

Observing that the latter expression defines a disjoint union of sets, by the countable disjoint additivity of measure, we have that

$$\mu(\cup_{i=1}^{\infty}(E_i \cap (\cup_{k=1}^{i-1} E_k)^c)) = \sum_{i=1}^{\infty} \mu(E_i \cap (\cup_{k=1}^{i-1} E_k)^c)$$

Now from the monotonicity of measure, and the fact that $\forall i$ $(E_i \cap (\cap_{k=1}^{i-1} E_k^c)) \subseteq E_i$, we have that $\mu(E_i \cap (\cap_{k=1}^{i-1} E_k^c)) \leq \mu(E_i)$.
Hence we have that

$$\sum_{i=1}^{\infty} \mu(E_i \cap (\cap_{k=1}^{i-1} E_k^c)) \leq \sum_{i=1}^{\infty} \mu(E_i)$$

Thus we have that

$$\mu(\bigcup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} \mu(E_i)$$

■

It turns out a number of the results that we have for Lebesgue measurable sets are going to go through as well.

Upwards Monotone Convergence: *Let $\{E_i\}$ be a sequence of $\mathcal{X}-$measurable sets with $E_1 \subseteq E_2 \subseteq \dots$. Then we have that*

$$\mu(\bigcup_{i=1}^{\infty} E_i) = \lim_{i\to\infty} \mu(E_i) = \sup_i \mu(E_i)$$

Proof. Begin by observing that if $\mu(E_i) = \infty$ for some $i \in \mathbb{N}$, we are done.
Now suppose that $\mu(E_i)$ is finite for all $i \in \mathbb{N}$. Now observe that $\cup_{i=1}^{\infty} E_i = \cup_{i=1}^{\infty}(E_i \setminus \cup_{k=1}^{i-1} E_k)$.
By the definition of $\{E_i\}$, we have that $\cup_{k=1}^{i-1} E_k = E_{i-1}$
Hence we have that $\cup_{i=1}^{\infty}(E_i \setminus \cup_{k=1}^{i-1} E_k) = \cup_{i=1}^{\infty}(E_i \setminus E_{i-1})$.
Now we simply apply μ to recover

$$\mu(\bigcup_{i=1}^{\infty} E_i) = \mu(\bigcup_{i=1}^{\infty} E_i \setminus E_{i-1}) = \sum_{i=1}^{\infty} \mu(E_i \setminus E_{i-1}) = \sum_{i=1}^{\infty} \mu(E_i) - \mu(E_{i-1})$$

Observing that this is a telescoping series[a], we recover that

$$\mu(\bigcup_{i=1}^{\infty} E_i) = \lim_{i\to\infty} \mu(E_i)$$

■

[a] Technically we should separate the sum into E_1 and everything else, but for the sake of brevity, just go with the slight algebra mistake, since it doesn't actually matter. Alternatively imagine that the sequence of sets has been prepended by the empty set.

Downwards Monotone Convergence: *Let $\{E_i\}$ be a sequence of $\mathcal{X}-$measurable sets with $\mathbb{R}^n \supseteq E_1 \supseteq E_2 \supseteq \dots$, and further suppose that $\mu(E_i) < \infty$ for some $i \in \mathbb{N}$. Then we have that*

$$\mu(\bigcap_{i=1}^{\infty} E_i) = \lim_{i\to\infty} \mu(E_i) = \inf_i \mu(E_i)$$

Proof. Observe that if $\mu(E_i)$ is finite for some $i \in \mathbb{N}$, then it must be finite in the limit. Without loss, consider only the truncation of this sequence past the first E_i with finite measure.
Observe that

$$E_1 \setminus \bigcap_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} E_i \setminus E_{i+1}$$

Applying $\mu(\cdot)$ to each side, we recover that

$$\mu(E_1 \setminus \bigcap_{i=1}^{\infty} E_i) = \mu(\bigcup_{i=2}^{\infty} E_i \setminus E_{i+1})$$

$$\mu(E_1) - \mu(\bigcap_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} (\mu(E_i) - \mu(E_{i+1}))$$

Observing that the right hand infinite scum contains a telescoping infinite series, we have that

$$\mu(E_1) - \mu(\bigcap_{i=1}^{\infty} E_i) = \mu(E_1) - \lim_{i\to\infty} \mu(E_{i+1})$$

Canceling $\mu(E_1)$, which is finite by assumption above, we have that

$$\mu(\bigcap_{i=1}^{\infty} E_i) = \lim_{i\to\infty} \mu(E_{i+1})$$

■

Dominated Convergence: *Let $\{E_i\}$ be a sequence of $\mathcal{X}$-measurable sets that converge to a set E, and further suppose that $E_i \subseteq F$ $\forall i \in \mathbb{N}$ with $\mu(F) < \infty$. Then E is measurable and $\lim_{i\to\infty} \mu(E_i) = \mu(E)$.*

Proof. Let us begin by showing measurability. To do so, observe that $\{E_i\} \to E$ if and only if

$$\limsup_{i\to\infty} E_i = \cap_{i=1}^{\infty}(\cup_{k=i}^{\infty} E_k) = \liminf_{i\to\infty} E_i = \cup_{i=1}^{\infty}(\cap_{k=i}^{\infty} E_k) = E$$

Now observe that since each E_i is measurable, and the countable union (and intersection) of measurable sets is measurable, we have that E is measurable.
Now observe that

$$\mu(E) = \mu(\cap_{i=1}^{\infty}(\cup_{k=i}^{\infty} E_k) = \lim_{i\to\infty} \mu(\cup_{k=i}^{\infty} E_k) \geq \lim_{i\to\infty} \mu(E_i)$$

Where the first equality follows from the convergence of $\{E_i\}$, the second equality follows from downwards monotone convergence, and the inequality follows from the fact that $E_i \subseteq \cup_{k=i}^{\infty} E_k$ and the monotonicity of measure.
Mirroring this, we observe that

$$\mu(E) = \mu(\cup_{i=1}^{\infty}(\cap_{k=i}^{\infty} E_k)) = \lim_{i\to\infty} \mu(\cap_{k=i}^{\infty} E_k) \leq \lim_{i\to\infty} \mu(E_i)$$

Where the first equality follows from the convergence of $\{E_i\}$, the second equality follows from upwards monotone convergence, and the inequality follows from the fact that $\cap_{k=i}^{\infty} E_k \subseteq E_i$ and the monotonicity of measure.
Combining inequalities, we have that

$$\mu(E) = \lim_{i\to\infty} \mu(E_i)$$

■

Dominated Convergence (Alternative Proof): *Let $\{E_i\}$ be a sequence of $\mathcal{X}$-measurable sets that converge to a set E, and further suppose that $E_i \subseteq F$ $\forall i \in \mathbb{N}$ with $\mu(F) < \infty$. Then E is measurable and $\lim_{i\to\infty} \mu(E_i) = \mu(E)$.*

Proof. Observe that $\{E_n\} \to E$ if and only if $\bigcap_{i=1}^{\infty} \bigcup_{k=i}^{\infty} (E_k \Delta E) = \emptyset$. Observing that

$$0 = \mu(\bigcap_{i=1}^{\infty} \bigcup_{k=i}^{\infty} (E_k \Delta E)) = \lim_{i\to\infty} \mu(\bigcup_{k=i}^{\infty} (E_k \Delta E)) \geq \lim_{i\to\infty} \mu(E_i \Delta E)$$

Recalling that by the definition of symmetric difference $E_i \Delta E = E_i \setminus E \cup E \setminus E_i$

$$\lim_{i\to\infty} \mu(E_i \setminus E \cup E \setminus E_i) \geq \lim_{i\to\infty} \mu(E_i \setminus E) = \lim_{i\to\infty} \mu(E_i) - \mu(E)$$

So we have that $\mu(E) \geq \lim_{i\to\infty} \mu(E_i)$.
Similarly we observe that

$$\lim_{i\to\infty} \mu(E_i \setminus E \cup E \setminus E_i) \geq \lim_{i\to\infty} \mu(E \setminus E_i) = \mu(E) - \lim_{i\to\infty} \mu(E_i)$$

So we have that $\lim_{i\to\infty} \mu(E_i) \geq \mu(E)$.
Hence we recover that

$$\lim_{i\to\infty} \mu(E_i) = \mu(E)$$

∎

In addition we have another important result in the form of the Borel-Cantelli Lemma.

Borel-Cantelli Lemma: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{E_i\}$ be a sequence of $\mathcal{X}$-measurable sets satisfying*

$$\sum_{i=1}^{\infty} \mu(E_i) < \infty$$

It follows then that

$$\mu(\limsup_{i\to\infty} E_i) = 0$$

Proof. Recall by the definition of the limit superior of a set,

$$\mu(\limsup_{i\to\infty} E_i) = \mu(\cap_{i=1}^{\infty}(\cup_{k=i}^{\infty} E_k))$$

Now observe by the monotonicity of measure, we have that

$$\mu(\cap_{i=1}^{\infty}(\cup_{k=i}^{\infty} E_k)) \leq \mu(\cup_{k=i}^{\infty} E_k)$$

Applying subadditivity, we have that

$$\mu(\cup_{k=i}^{\infty} E_k) \leq \sum_{k=i}^{\infty} \mu(E_k)$$

Observing that the convergence of the infinite series implies that the tail goes to zero, we have that

$$\lim_{i\to\infty} \sum_{k=i}^{\infty} \mu(E_k) = 0$$

As a consequence we have that

$$\mu(\limsup_{i\to\infty} E_i) \leq 0$$

Which, together with the non-negativity of measure, implies that

$$\mu(\limsup_{i\to\infty} E_i) = 0$$

■

We say that a measure space $(X, \mathcal{X}, \mu)$ is **complete** if it has the property that

$$E \in \mathcal{X} \wedge \mu(E) = 0 \implies \forall F \subseteq E \; F \in \mathcal{X}$$

In other words, we say a measure space is complete if it has the property that every sub-null set is measurable.

As was mentioned before, it turns out that the Borel σ-algebra and the Lebesgue σ-algebra are closely related to each other. Specifically, $(\mathbb{R}, \mathcal{L}, \lambda)$ is the completion of $(\mathbb{R}, \mathcal{B}, \lambda)$.[30]

[30] How do we know that the Borel sigma algebra is not complete? Consider that we know that the Cantor set has Lebesgue measure zero, but has uncountably many elements. As a consequence, it must have $2^{\aleph_1}$ possible subsets. One can show that the cardinality of the Borel $\sigma-$algebra is exactly that of the continuum, so clearly not every subset of the Cantor set can be measurable.

We generalize measurable functions in a somewhat more opaque way. Specifically, let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces. We say that a function $f : X \to Y$ is measurable if it has the property that $\forall F \in \mathcal{Y}\ f^{-1}(F) \in \mathcal{X}$. Again, the analogy to continuous functions is clear. A measurable function is simply a structure preserving map from one measurable space to another. What is the structure preserved? Measurability. This is in contrast to continuous functions, which are also structure preserving maps, but preserve topological properties of the spaces, as opposed to measurability.

Measurable functions viewed in this way also has the advantage of seeing exactly why the composition of Lebesgue-measurable functions need not be measurable, but the composition of Borel-measurable functions is. It is this latter class of functions that we are more typically concerned with, for both this reason and for others.

Just as we constructed simple integrals and unsigned integrals in the context of the Lebesgue measure, we can do basically the same thing (with a bit more notational machinery) to construct simple integrals and unsigned integrals on measure spaces writ large.

Specifically, let $(X, \mathcal{X}, \mu)$ be a measure space, and define a simple function as a measurable function that takes on finitely many values, $a_1, \ldots, a_k$. This gives us the following definition of a simple integral.

$$Simp \int_X f d\mu \equiv \sum_{i=1}^{k} a_i \mu(\{f^{-1}(a_i)\})$$

[31]

One particularly important property of simple integrals is that they are insensitive to changes on a null-set. Another important property is that they exhibit linearity.

[31] Recall again that we can write any simple function in terms of disjoint indicator functions. This is directly equivalent to just pulling back the indicator function and defining it in terms of the value it takes on that set in $[0, \infty]$.

Linearity of the Simple Integral: *Suppose that f and g are two simple functions from a measure space $(X, \mathcal{X}, \mu)$ to $\mathbb{R}$. Then we have that:*

1. $c Simp \int_X f d\mu = Simp \int_X cf d\mu \; \forall c \in \mathbb{R}_+$
2. $Simp \int_X f + g d\mu = Simp \int_X f d\mu + Simp \int_X g d\mu$

Proof. Let our assumptions be as above, and observe that

$$c Simp \int_X f d\mu = c \sum_{i=1}^{k} a_i \mu(\{f^{-1}(a_i)\}) = \sum_{i=1}^{k} c a_i \mu(\{f^{-1}(a_i)\}) = Simp \int_X cf d\mu$$

Where the first equality follows from the fact that f is simple, the second from the linearity of summation, and the third from the fact that a scalar multiple of f remains a simple function.
For additivity, the argument is not hard but it is tedious. Basically we pursue a Venn diagram argument. Assume without loss that both f and g have a standard representation. Observe that we can write

$$Simp \int_X f d\mu = \sum_{i=1}^{k} a_i \mu(\{f^{-1}(a_i)\})$$

and

$$Simp \int_X g d\mu = \sum_{i=1}^{k'} b_i \mu(\{g^{-1}(b_i)\})$$

There are most $2^{k+k'}$ sets in play here. Assume without loss that this is the number of sets in play, calling it k''.
Then we can write the above expressions as

$$\sum_{i=1}^{k} a_i \mu(\{f^{-1}(a_i)\}) = \sum_{i=1}^{k''} a_i \mu(E_i)$$

$$\sum_{i=1}^{k'} b_i \mu(\{g^{-1}(b_i)\}) = \sum_{i=1}^{k''} b_i \mu(E_i)$$

[a]

Here applying linearity and grouping terms we have that

$$Simp \int_X f d\mu + Simp \int_X g d\mu = \sum_{i=1}^{k''} a_i \mu(E_i) + \sum_{i=1}^{k''} b_i \mu(E_i) = \sum_{i=1}^{k''} (a_i + b_i) \mu(E_i) = Simp \int_X f + g d\mu$$

■

[a] Note that here we are not representing the simple function in a minimal representation. Specifically, the coefficients may repeat over time.

Likewise, we can define an unsigned integral by

$$\int_X f d\mu \equiv \sup_{0\leq g\leq f;\text{ g simple}} Simp \int_X g d\mu$$

This integral has all the typical nice properties that we expect from integrals.

Claim: *Let f, g be measurable functions from X to $[0, \infty)$. If $\mu(\{x \in X | f < g\}) = 0$, then $\int_X f d\mu \geq \int_X g d\mu$.*

Proof. Let f, g be given. Observe that any simple function h has a simple integral exactly equal to a simple function h' where $\mu(\{h' \neq h\}) = 0$. Since $f \geq g$ μ-almost everywhere, if a simple function r lies below g, then we can modify r on a null set so that it lies below f. Hence

$$\int_X f d\mu \geq \int_X g d\mu$$

■

Claim: *Let f be a measurable functions from X to $[0, \infty)$. Then $\forall c \in \mathbb{R}_+$ we have that*

$$\int_X c f d\mu = c \int_X f d\mu$$

Proof. This is somewhat tricky. The cleanest way to do it is to use the monotone convergence theorem (see below) on a sequence of simple functions that converges to f.

■

Claim: *Let f, g be measurable functions from X to $[0, \infty)$. Then $\int_X f d\mu + \int_X g d\mu \leq \int_X f + g d\mu$.*

Proof. Let f, g be given. Consider the set of simple functions $\underset{\sim}{F} = \{h | 0 \leq h \leq f \text{ with h simple}\}$, and $\underset{\sim}{G} = \{w | 0 \leq w \leq g \text{ with w simple}\}$. Pick $h \in \underset{\sim}{F}$ and $w \in \underset{\sim}{G}$ and observe that the $h + w$ must lie below $f + g$. Hence we have that[a]

$$\int_X f d\mu + \int_X g d\mu \leq \int_X f + g d\mu$$

■

[a]Observe if $h + w$ is simple and lies below $f + g$, it can be at most as close to $f + g$ as the supremum of simple functions lying below $f + g$.

One can also establish superadditivity, but the proof is a bit complicated. Once one has superadditivity though, we recover the typical separation property of integrals, namely that

$$\int_X f d\mu + \int_X g d\mu = \int_X f + g d\mu$$

And as was done with the Lebesgue integral, we can also define in generality an **absolutely convergent integral**.

Let $(X, \mathcal{X}, \mu)$ be a measure space, and $f : X \to \mathbb{R}$ a measurable function. We say f is absolutely integrable if

$$||f||_{L_1(X,\mathcal{X},\mu)} \equiv \int_X |f| d\mu < \infty$$

As with the e.g. the Riemann integral, we define the integral of a signed function (to put this in in a very loose and nontechnical sense) as the area of the curve above $y = 0$ less the area below. More formally, we write that:

$$\int_X f d\mu \equiv \int_X f_+ d\mu + \int_X f_- d\mu$$

Where $f_+ \equiv \max(f, 0)$ and $f_- \equiv \max(-f, 0)$

There are a number of results that we have for unsigned integrals that we would like to generalize to absolutely convergent integrals. The two[32] linearity and monotonicity. I prove linearity first.

[32] Or really three if you consider linearity as two separate ones.

Linearity of the Absolutely Convergent Integral: *Suppose that f and g are two absolutely integrable functions from a measure space $(X, \mathcal{X}, \mu)$ to $\mathbb{R}$. Then we have that:*

1. $c\int_X f d\mu = \int_X cf d\mu \forall c \in \mathbb{R}_+$
2. $\int_X f + g d\mu = \int_X f d\mu + \int_X g d\mu$

Proof. Let our assumptions be as above, and observe that

$$\int_X cf d\mu = \int_X cf_+ d\mu - \int_X cf_- d\mu = c\int_X f_+ d\mu - c\int_X f_- d\mu = c\int_X f d\mu$$

Where the first equality follows from the definition of the absolutely convergent integral, the second from the linearity of the unsigned integral, and the third again from the definition of the absolutely convergent integral.
Now we prove additivity.
Let $h = f + g$, and observe that

$$h_+ - h_- = f_+ + g_+ - h_- - g_-$$

[a]

Doing algebra, one recovers that

$$h_+ + f_- + g_- = h_- + f_+ + g_+$$

Integrating each side and applying linearity of the unsigned integral, we have that

$$\int_X h_+ d\mu + \int_X f_- d\mu + \int_X g_- d\mu = \int_X h_- d\mu + \int_X f_+ d\mu + \int_X g_+ d\mu$$

Again doing algebra, we recover that

$$\int_X h_+ d\mu - \int_X h_- d\mu = \int_X f_+ d\mu - \int_X f_- d\mu + \int_X g_+ d\mu - \int_X g_- d\mu$$

Substituting again for the definition of the absolutely convergent integral, we have that

$$\int_X h d\mu = \int_X f + g d\mu = \int_X f d\mu + \int_X g d\mu$$

Which was what was wanted. ■

[a] Note that it is not in general true that $h_+ = f_+ + g_+$. To see this, think of say the sum of cosine and sine, or really any two continuous functions which switch signs at different points.

Monotonicity of the Absolutely Convergent Integral: *Suppose that f and g are two absolutely integrable functions from a measure space $(X, \mathcal{X}, \mu)$ to $\mathbb{R}$. Suppose that f pointwise dominates g μ-almost everywhere. Then we have that:*

$$\int_X f d\mu \geq \int_X g d\mu$$

Proof. Let the assumptions be as above. Without loss of generality we can modify f on a null set so that it dominates g everywhere. Then we have that

$$\int_X f_+ d\mu \geq \int_X g_+ d\mu \wedge \int_X g_- d\mu \geq \int_X f_- d\mu$$

[a]

Hence we have that

$$\int_X f_+ d\mu - \int_X f_- d\mu \geq \int_X g_+ d\mu - \int_X g_- d\mu$$

Which after applying the definition of the absolutely convergent integral delivers us the result. ∎

[a] To see this observe that if $f \geq g$, then $\mu(\{x \in X | f \geq 0\}) \geq \mu(\{x \in X | f \geq 0\})$. Coupling this with the observation that f is always weakly more positive, or weakly less negative, and the monotonicity of the unsigned integral we have the claimed inequalities.

Similarly to what we did for sets, we also have a number of potent convergence theorems for integrals here.

Monotone Convergence Theorem: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $0 \leq f_1 \leq f_2 \leq \dots$ be a monotone non-decreasing sequence of unsigned measurable functions on X. Then we have*

$$\lim_{i \to \infty} \int_X f_i d\mu = \int_X \lim_{i \to \infty} f_i d\mu$$

Proof. Let $f \equiv \lim_{i \to \infty} f_i = \sup_i f_i$, where the equality follows from the fact that $\{f_i\}$ is a non-decreasing sequence of functions.

By definition of the supremum and the monotonicity of the unsigned integral, we have that

$$\int_X f_i d\mu \leq \int_X f d\mu = \int_X \lim_{i \to \infty} f_i d\mu$$

Taking limits and observing that weak inequalities are preserved under limits, we have that

$$\lim_{i \to \infty} \int_X f_i d\mu \leq \int_X f d\mu = \int_X \lim_{i \to \infty} f_i d\mu$$

It remains to demonstrate that

$$\int_X f d\mu \leq \lim_{i \to \infty} \int_X f_i d\mu$$

Begin by recalling that by definition, $\int_X f d\mu = \sup_{0\leq g\leq f \text{ with } g \text{ simple}} Simp \int_X g d\mu$
Hence, it suffices to show that

$$\int_X g d\mu \leq \lim_{i\to\infty} \int_X f_i d\mu$$

For any simple g that is pointwise dominated by f.
Let us begin by assuming without loss that g is finite everywhere[a].
Since g is simple, we can write

$$\int_X g d\mu = \sum_{j=1}^{k} c_j \mu(A_j), c_j \in \mathbb{R}_+, A_j \cap A_l = \emptyset \; \forall j \neq l$$

Now let $\epsilon > 0$ be given. Observe that by definition of f, we have

$$f(x) = \sup_i f_i(x) > (1-\epsilon)c_j \forall x \in A_j$$

Defining $A_{j,i} = \{x \in A_j | f_i(x) > (1-\epsilon)c_j\}$, we observe that $A_{j,i}$ is measurable $\forall j, i$.[b]
Further, since $\{f_i\}$ is a monotonic, non-decreasing sequence of functions, we also have that $A_{j,i} \subseteq A_{j,k} \; \forall i < k$
By the monotone convergence of μ-measurable sets, we have that $\lim_{i\to\infty} \mu(A_{j,i}) = \mu(A_j)$.
Now observe that $f_i \geq \sum_{j=1}^{k} (1-\epsilon) c_j 1_{A_{j,i}}$.[c]
Integrating this pointwise limit delivers that

$$\int_X f_i d\mu \geq \sum_{j=1}^{k} (1-\epsilon) c_j \mu(A_{j,i})$$

Finally, taking limits, we have that

$$\lim_{i\to\infty} \int_X f_i d\mu \geq \sum_{j=1}^{k} (1-\epsilon) c_j \mu(A_j)$$

Since ϵ was arbitrary, we thus have that

$$\int_X f d\mu \leq \lim_{i\to\infty} \int_X f_i d\mu$$

Combining inequalities then, we have that

$$\lim_{i\to\infty} \int_X f_i d\mu = \int_X \lim_{i\to\infty} f_i d\mu$$

Which was what was wanted. ■

[a] We can do this because of what is called the vertical truncation property of the unsigned integral, i.e. that $\lim_{n\to\infty} \int_X min(f,n) d\mu = \int_X f d\mu$

[b] This is because the A_j are assumed measurable sets, and by definition of measurability - given that we are concerned with Borel sets in the range - the inverse image of values of measurable functions on some open set in the range must be also be measurable. As the intersection of measurable sets is itself measurable, we thus have that $A_{j,i}$ is measurable.

[c] To see why this is true, observe that the construction of $A_{j,i}$ is exactly those sets of points that make this true. Further observe that by the definition of the indicator function, all points outside that set are mapped to 0. Since by assumption $f_i \geq 0 \forall i$, we have the above inequality holds.

Fatou's Lemma: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{f_i\}$ be a sequence of unsigned measurable functions. Then*

$$\int_X \liminf_{i\to\infty} f_n d\mu \leq \liminf_{i\to\infty} \int_X f_i d\mu$$

Proof. Begin by recalling that $\liminf_{i\to\infty} f_i = \lim_{i\to\infty} \inf_{k>i} f_k = \sup_i \inf_{k\geq i} f_k$.
Now observe that $\inf_{k\geq i} f_k$ defines a monotonically non-decreasing sequence of measurable functions in k. From the monotone convergence theorem, then, we have that

$$\sup_i \int_X \inf_{k\geq i} f_k d\mu = \int_X \sup_i \inf_{k\geq i} f_k d\mu$$

.
Now let us observe that $\inf_{k\geq i} \leq f_i$. By the monotinicity of the unsigned integral, then, we have that

$$\int_X \inf_{k\geq i} f_k d\mu \leq \int_X f_k d\mu \; \forall k \geq i$$

It then follows that

$$\int_X \inf_{k\geq i} f_k d\mu \leq \inf_{k\geq i} \int_X f_k d\mu$$

Applying the monotone convergence result, we have

$$\int_X \sup_i \inf_{k\geq i} f_k d\mu = \sup_i \int_X \inf_{k\geq i} f_k d\mu \leq \sup_i \inf_{k\geq i} \int_X f_k d\mu$$

Substituting for definitions, we recover:

$$\int_X \liminf_{i\to\infty} f_i d\mu \leq \liminf_{i\to\infty} \int_X f_i d\mu$$

∎

One can also see that Fatou's Lemma provides us with a slightly more general form of the Monotone Convergence Theorem.

Convergence via Pointwise Domination by Limiting Function: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{f_i\}$ be a sequence of unsigned $\mathcal{X}$-measurable functions such that $f_i \to f$, and $f_i \leq f$ $\forall i$. Then we have that*

$$\lim_{i\to\infty} \int_X f_i d\mu = \int_X f d\mu$$

Proof. Observe that by the pointwise domination of f_n by f, and the monotonicity of the unsigned integral, we have that

$$\int_X f_i d\mu \leq \int_X f d\mu$$

By definition of the limit superior and limit inferior, we have that

$$\liminf_{i\to\infty} \int_X f_i d\mu \leq \limsup_{i\to\infty} \int_X f_i d\mu \leq \int_X f d\mu$$

Applying Fatou's Lemma and observing the convergence of f_i to f implies $\liminf_{i\to\infty} f_i = f$, we have that

$$\int_X f d\mu \leq \liminf_{i\to\infty} \int_X f_i d\mu \leq \limsup_{i\to\infty} \int_X f_i d\mu \leq \int_X f d\mu$$

By the Squeeze Theorem then, we have

$$\lim_{i\to\infty} \int_X f_i d\mu = \int_X f d\mu$$

Which was what was wanted. ■

Dominated Convergence Theorem: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{f_i\}$ be a sequence of measurable functions from X to $\mathbb{R}$ that converges pointwise μ-almost everywhere to a measurable function $f : X \to \mathbb{R}$. Further suppose that $\forall i$ $|f_i|$ is pointwise μ-almost everywhere dominated by an unsigned absolutely integrable function g. Then we have that*

$$\lim_{i\to\infty}\int_X f_i d\mu = \int_X f d\mu$$

Proof. Assume as above. By modifying f_i and/or f on a null set we can assume without loss of generality that $\{f_i\} \to f$ pointwise everywhere and further that f_i is bounded everywhere by g.
Hence we have that
$-g \leq f_i \leq g$
Now we apply Fatou's Lemma to $f_i + g$ to recover that

$$\int_X f + g d\mu \leq \liminf_{i\to\infty}\int_X f_i + g d\mu$$

By the linearity of the unsigned integral, and the absolute integrability of g, we thus have that

$$\int_X f d\mu \leq \liminf_{i\to\infty}\int_X f_i d\mu$$

Similarly, we apply Fatou's Lemma to $G - f_i$ to recover that

$$\int_X g - f d\mu \leq \liminf_{i\to\infty}\int_X g - f_i d\mu$$

Repeating the same step, we have that

$$\int_X -f d\mu \leq \liminf_{i\to\infty}\int_X -f_i d\mu$$

Clearly this is equivalent to

$$\int_X f d\mu \geq \limsup_{i\to\infty}\int_X f_i$$

Since by definition of limit superior and limit inferior we have that

$$\liminf_{i\to\infty}\int_X f_i d\mu \leq \limsup_{i\to\infty}\int_X f_i d\mu$$

The result is shown. ∎

2.6 Probability Measures

Modern probability theory tends to use a different set of notation for the same underlying mathematical formalism. So, for instance, where we have prior referred to $(X, \mathcal{X}, \mu)$ as a measurable space, a probabilist will often refer to the same structure as a **probability triple** $(\Omega, \mathcal{F}, \mathbb{P})$. This is just a renaming of the underlying structure and so has embedded in it the same mathematical formalism, but it has a slightly different interpretation. Ω is what is referred to as the **sample space**: it consists of the set of possible outcomes. $\mathcal{F}$ is what is referred to as an **event space**, and $\mathbb{P}$ is of course a probability measure, as was defined in the prior section.

To make this concrete, consider the roll of a six-sided die. Assuming the die is fair, there are six possible outcomes: the die comes up with 1, with 2, with 3, and so on.

Each of those possible outcomes is essentially a state in the sample space.

If we observe the outcome of the die roll, then we know exactly where in the sample space we are. In this case, the σ-algebra that represents the die roll is precisely the discrete σ-algebra, and the probability measure we want to use is the obvious one.

Suppose, however, that one can cover the sides of the die with pieces of blue paper or red paper. Suppose as well that one chooses to cover 1,3, and 5 with red paper, and 2,4, and 6 with blue. Then, what we observe is not the true "state" of the system (i.e. the number that we landed on), but rather the event of landing on blue or red paper. This partition of the sample space naturally gives rise to a different sigma algebra.

In this case, $\mathcal{F} = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \{1,2,3,4,5,6\}\}$, and the probability measure is the obvious one.

Keeping this interpretation in mind, we now turn to random variables. Specifically, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple, and let $X : \Omega \to B$, where B is equipped with some σ-algebra $\mathcal{B}$.[33] If f is measurable, we say that f is a **random variable**. Further, we write that the probability of X taking on some value in a set $E \subseteq B$ is given by

$$P(X \in E) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in E\})$$

Now, one might notice that when we are, say, calculating expectations, we do not typically reference the probability measure on the sample space we are interested in, but rather the cumulative distribution function for the continuous (or discrete) random variable we are interested in. To perhaps put this conundrum concretely, let X be a random variable from some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathbb{R}$ equipped with the Borel σ-algebra. We don't typically write that

$$\mathbb{E}[g(X)] = \int_\Omega g \circ X d\mathbb{P}$$

Rather we write that

$$\mathbb{E}[g(X)] = \int_\mathbb{R} g(x) dF(x)$$

[34]

Are these two things the same? To answer this question requires that we define a **Push-Forward Measure**:

Let $(X, \mathcal{X}, \mu)$ be a measure space, and let f be a measurable function mapping X to some measurable space, $(Y, \mathcal{Y})$. We say that the push-forward of μ is given by

$$f_*(\mu)(E) = \mu(f^{-1}(E)) \forall\, E \in \mathcal{Y}$$

It remains to show that the push-forward measure does in fact satisfy the measure axioms, but in order to show this we must first prove a lemma, namely that set unions are nicely behaved with respect to inverse images.

[33] A particularly common choice is $\mathbb{R}$ equipped with the Borel σ-algebra.

[34] This expression is what is referred to as the Law of the Unconscious Statistician (LOTUS).

Lemma: *Let E_i be a collection of sets. Then*

$$f^{-1}(\bigcup_{i=1}^{\infty} E_i) = \bigcup_{i=1}^{\infty} f^{-1}(E_i)$$

Proof. Let $x \in f^{-1}(\bigcup_{i=1}^{\infty} E_i)$. Then $f(x) \in \bigcup_{i=1}^{\infty} E_i$. Hence $\exists j \in \mathbb{N} | f(x) \in E_j$. Then $x \in f^{-1}(E_j)$, which further implies that $x \in \bigcup_{i=1}^{\infty} f^{-1}(E_i)$. The other case is symmetric. ∎

Claim: *A push-forward measure $f_*(\mu)(\cdot)$ satisfies the measure axioms.*

Proof. Let us begin by demonstrating $f_*(\mu)(\cdot)$ satisfies non-negativity. Let E be an arbitrary subset of the codomain of f, and observe that $f_*(\mu)(E) = \mu(f^{-1}(E))$. Since f is measurable, $f^{-1}(E)$ is a measurable set in $(X, \mathcal{X}, \mu)$ which ensures that $\mu(f^{-1}(E)$ both exists and is contained in $[0, \infty]$. Hence non-negativity is satisfied.
Now consider the empty set axiom. So let us consider $f_*(\mu)(\emptyset)$. By the definition of a push-forward measure $f_*(\mu)(\emptyset) = \mu(f^{-1}(\emptyset)$. Since the preimage of the empty set is empty, we have that $\mu(f^{-1}(\emptyset) = \mu(\emptyset) = 0$.
Now we wish to show countable disjoint additivity. Let $\{E_i\}_{i=1}^{\infty}$ be a countable set of disjoint subsets E_i of Y.
Observe that

$$f_*(\mu)(\bigcup_{i=1}^{\infty} E_i) = \mu(f^{-1}(\bigcup_{i=1}^{\infty} E_i)) = \mu(\bigcup_{i=1}^{\infty} f^{-1}(E_i))$$

Where the first equality follows from the definition of the push-forward measure, and the second from the lemma.
Now, observing that $f^{-1}(E_i) \cap f^{-1}(E_j) = \emptyset^a$, we thus have that

$$\mu(\bigcup_{i=1}^{\infty} f^{-1}(E_i)) = \sum_{i=1}^{\infty} \mu(f^{-1}(E_i)) = \sum_{i=1}^{\infty} f_*(\mu)(E_i)$$

Where the first equality follows from the disjoint nature of the inverse images, and the second equality follows from the definition of a push-forward measure.
Hence we have that a push-forward measure is, in fact, a measure. ∎

[a] Suppose not. Then there exists i, j such that $f^{-1}(E_i) \cap f^{-1}(E_j) \neq \emptyset$. Pick x in that intersection, and observe that by the definition of the inverse image $f(x) \in E_i$ and $f(x) \in E_j$. But $E_i capE_j = \emptyset$ by assumption, so then f is no longer single valued at x, which is impossible.

In the context of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we call the push-forward measure induced by a random variable X the **Distribution Measure of** X. If we are concerned with a random variable that maps to $(\mathbb{R}, \mathcal{B})$, then this is equivalent to the cumulative distribution function of the random variable X.[35]

We need one more result to get the equivalence between the two expectations above, namely the following.

[35] This is because $\mathcal{B}$ is generated by the set of half-open intervals of the form $(-\infty, a]$ $\forall a \in \mathbb{R}$.

Change of Variables: *Let $(X, \mathcal{X}, \mu)$ be a measure space, $(Y, \mathcal{Y})$ a measurable space, and $\phi : X \to Y$ a measurable function. Additionally, let $f : Y \to [0, \infty]$ be a $\mathcal{Y}$-measurable function. Denoting the push forward measure induced by ϕ as $\phi_*(\mu)(\cdot)$, we then have that:*

$$\int_X (f \circ \phi) d\mu = \int_Y f d\phi_*(\mu)$$

Proof. We begin by supposing that f is simple. Then $f = \sum_{i=1}^k a_i 1_{E_i}$, where $E_i = \{y \in Y | f(y) = a_i\}$. Observing that

$$\int_Y f d\phi_*(\mu) = \sum_{i=1}^k a_i \phi_*(\mu)(f^{-1}(\{a_i\})) = \sum_{i=1}^k a_i \mu(\phi^{-1}(f^{-1}(\{a_i\})) = \int_X (f \circ \phi) d\mu$$

We have the result for simple functions.
Now, we wish to show that $f \circ \phi$ is a measurable function. Let $E \subseteq [0, \infty]$ be a $\mathcal{B}$-measurable set. By the $\mathcal{Y}$-measurability of f, we have that $f^{-1}(E)$ is a measurable set in Y. Now by the $\mathcal{X}$-measurability of ϕ, we have that $\phi^{-1}(f^{-1}(E))$ is a measurable set. Hence $f \circ \phi$ is an $\mathcal{X}$-measurable function.
Now observe that the fact that f is $\mathcal{Y}$-measurable implies that there exists a sequence of simple functions pointwise dominated by f that converge to f. Calling this sequence of functions $\{h_i\}$, we have, by *Convergence via Pointwise Domination by Limiting Function* result that

$$\lim_{i \to \infty} \int_Y h_i d\phi_*(\mu) = \int_Y f d\phi_*(\mu)$$

But observe from the prior argument that $\forall i$, we have that $\int_Y h_i d\phi_*(\mu) = \int_X (h_i \circ \phi) d\mu$ Hence we have that

$$\lim_{i \to \infty} \int_X (h_i \circ \phi) d\mu = \lim_{i \to \infty} \int_Y h_i d\phi_*(\mu)$$

Applying the *Convergence via Pointwise Domination by Limiting Function* result again, we have that

$$\lim_{i \to \infty} \int_X (h_i \circ \phi) d\mu = \int_X (f \circ \phi) d\mu$$

Combining equalities, we then have that

$$\int_X (f \circ \phi) d\mu = \int_Y f d\phi_*(\mu)$$

Which was what was wanted. ■

A couple of observations here. The first is that this implies that we consider distributions of random variables without loss, because value through integrals is preserved under measurable transformations. The second is that we will always have a cumulative distribution function induced by a random variable, because all that is required for a push-forward measure is some measurable function, and a random variable is by definition a $\mathcal{F}$-measurable function.

The natural next question is of course, what about densities. Will every continuous (resp.

discrete) random variable also induce a probability density (resp. mass) function? The answer is no.[36] In general, the measure-theoretic foundations of probability density functions are (considerably) more complicated than have been presented here, but their existence arises as a result of the Radon-Nikodym Theorem.

2.7 Modes of Convergence

When you are dealing with, say, a sequence of real numbers, there is no ambiguity as to what it means for a sequence to converge. Specifically, let $\{x_i\}$ be a sequence of real numbers. We say that $x_i \to x$ if $\forall \epsilon > 0\ \exists N_\epsilon\ |\forall i \geq N|x_i - x| < \epsilon$. This definition broadly survives generalization to $\mathbb{R}^n$, where we say that $x_i \to x$ if $\forall \epsilon > 0\ \exists N_\epsilon\ \forall i > N_\epsilon ||x_i - x|| < \epsilon$. Here $|| \cdot ||$ can be chosen to be some arbitrary norm on $\mathbb{R}^n$.

Functions obviously are a different beast because they are not defined at a single point, and thus efforts to characterize their convergence are necessarily going to require some consideration of the entirety of the domain of the function.

We have discussed previously two modes of convergence, these being **pointwise convergence**:

$$f_i \to f \iff \forall \epsilon > 0 \forall x \in X \exists N_{\epsilon,x} | \forall i > N_{\epsilon,x} |f_i(x) - f(x)| < \epsilon$$

And uniform convergence

$$f_i \to f \iff \forall \epsilon > 0 \exists N_\epsilon | \forall x \in X \forall i > N_\epsilon |f_i(x) - f(x)| < \epsilon$$

But because we are interested in measurable functions f on a measure space $(X, \mathcal{X}, \mu)$, we can actually broaden this quite substantially.

Here are five that people typically are interested in.

1. **Pointwise almost everywhere convergence**: We say $f_i \to f$ if $f_i \to f$ pointwise μ-almost everywhere. Or to put it slightly differently, the set of points where f_i does not converge to f pointwise has measure zero with respect to μ.

2. **Uniformly almost everywhere convergence**:[37] We say $f_i \to f$ if $\forall \epsilon > 0\ \exists N | \forall i > N |f_i(x) - f(x)| < \epsilon$ μ-almost everywhere.

3. **Almost uniform convergence**: $f_i \to f$ if $\forall \epsilon > 0\ \exists \in \mathcal{X}$ with $\mu(E) \leq \epsilon$ and $f_i \to f$ on $X \setminus E$.

4. **Convergence in L_1 norm**: $f_i \to f$ if $||f_n - f||_{L_1(\mu)} = \int_X |f_n(x) - f(x)| d\mu \to 0$ as $i \to \infty$.

5. **Convergence in measure**: $f_i \to f$ if $\forall \epsilon > 0\ \mu(\{x \in X : |f_i(x) - f(x)| \geq \epsilon\}) \to 0$ as $i \to \infty$.

In the context of probability theory, these underlying concepts of course have slightly different names. Probabilists typically refer to L_1 convergence as convergence in mean. They refer to pointwise convergence μ-almost everywhere convergence as almost sure convergence. And they refer to convergence in measure as convergence in probability. These, like the prior convention to

[36] For an example, look up the Cantor Distribution.
[37] Also called $L^\infty(\mu)$ convergence.

refer to a specific kind of measure spaces as a probability space amounts to not much more than a renaming of the underlying formalism.[38]

Now we want to show that these modes of convergence are nicely behaved in certain manners.

Claim: *Let $(X, \mathcal{X}, \mu)$ be a measure space, $\{f_i\}$ a sequence of $\mathcal{X}$-measurable functions from X to $\mathbb{R}$, and $f : X \to \mathbb{R}$ be $\mathcal{X}$-measurable. Then if $\{f_i\} \to f$ in one sense, $\{|f_i - f|\} \to 0$ in the same sense.*

Proof. Observe that $\forall x \in X$, we have that $|f_i(x) - f(x)| = |f_i(x) - f(x)| - 0 = ||f_i(x) - f(x)| - 0|$. This observation delivers us the result. ∎

[38] In fact, whereas at least with a probability space, there is an additional restriction placed on the measure, here we do not have such an additional restriction apart from perhaps that these concepts are only talked about with respect to some underlying probability space.

Additivity of Convergence: *Suppose that $\{f_i\} \to f$ along one of the aforementioned modes of convergence, and that $\{g_i\} \to g$ along the same mode. Then $\{f_n + g_n\} \to f + g$, and $\{cf_i\} \to cf$ along the same mode.*

Proof. Let us begin with pointwise convergence.
Let $\epsilon > 0$ be given.
Observe that $\{f_i \to f\}$ implies that $\forall \epsilon_f > 0 \forall x \in X \exists N^f_{\epsilon,x}$ such that $\forall i \geq N^f_{\epsilon,x} |f_i(x) - f(x)| < \epsilon_f$. We recover the equivalent from $\{g_i \to g\}$. Letting $\epsilon_f = \epsilon_g = \frac{\epsilon}{2}$ and setting $N^{f+g}_{\epsilon,x}$ equal to the max of $N^f_{\epsilon,x}$, and $N^g_{\epsilon,x}$, we recover that for that for $i \geq N^{f+g}_{\epsilon,x}$, we have that $|f_i(x) - f(x)| < \frac{\epsilon}{2}$ and $|g_i(x) - g(x)| < \frac{\epsilon}{2}$. Adding the two, we recover that $|f_i(x) - f(x)| + |g_i(x) - g(x)| < \epsilon$.
Applying the triangle inequality and associativity of addition delivers us that

$$|f_i(x) + g_i(x) - f(x) - g(x)| < \epsilon$$

An almost identical proof will deliver the same result for uniform convergence, and the result for μ-almost everywhere convergence of both kinds follows from the observation that the measure of the union of two measure zero sets is at most 0.
The result for almost uniform convergence is delivered by picking $\epsilon_f = \epsilon_g = \frac{\epsilon}{2}$ and then simply applying the result for uniform convergence.
For the L_1 norm result, observe that $\forall \epsilon_f > 0$ we have that there exists N_{ϵ_f} such that $\forall i \geq N_{\epsilon_f}$ we have that $\int_X |f_i(x) - f(x)| d\mu < \epsilon_f$, and likewise for g. Picking $\epsilon_f = \epsilon_g = \frac{\epsilon}{2}$ and taking N^{f+g}_{ϵ} equal to the max of N_{ϵ_f} and N_{ϵ_g}, delivers us that $\int_X |f_i(x) - f(x)| d\mu + \int_X |g_i(x) - g(x)| d\mu < \epsilon$.
By the linearity of the unsigned integral, we have that

$$\int_X |f_i(x) - f(x)| d\mu + \int_X |g_i(x) - g(x)| d\mu = \int_X |f_i(x) - f(x)| + |g_i(x) - g(x)| d\mu$$

From the pointwise triangle inequality and the monotonicity of the unsigned integral, we then have that

$$\int_X |f_i + g_i - f - g| d\mu \leq \int_X |f_i(x) - f(x)| + |g_i(x) - g(x)| d\mu < \epsilon$$

This delivers the result.
Now for convergence in measure, we basically apply the same trick we did in the prior cases, just by picking N in such a way as to ensure that the two sets are sufficiently small. ■

Multiplicativity of Convergence: *Suppose that $\{f_i\} \to f$ along one of the aforementioned modes of convergence and that $c \in \mathbb{R}$. Then $\{cf_n\} \to cf$ along the same mode.*

Proof. As with the prior claim, we begin by considering pointwise convergence.
Let $\epsilon > 0$ be given. We can pick N_f such that $|f_i(x) - f(x)| < \frac{\epsilon}{|c|}$. This implies that $|c||f_i(x) - f(x)| < \epsilon$. Since $|p||q| = |pq|$, we have that

$$|cf_i(x) - cf(x)| < \epsilon$$

Delivering us the result.
As with before, almost identical arguments deliver us uniform convergence and μ-almost everywhere convergence of both kinds.
Then we pick $\epsilon_f = \epsilon_g = \frac{\epsilon}{2}$ and applying the uniform convergence result to recover the uniform convergence result.
To prove the L_1 norm result, observe that we have that by assumption that

$$||f_i - f||_{L_1(\mu)} \to 0$$

By the absolute homogeneity property of norms, we also have that $\forall c \in \mathbb{R}\ \forall i\ |c|||f_i - f||_{L_1(\mu)} = ||cf_i - cf||_{L_1(\mu)}$. This delivers us the result.
Finally, we prove the convergence in measure result.
Let $\epsilon > 0$ be given. Define $\epsilon' = \frac{\epsilon}{|c|}$. By the convergence of measure of $\{f_i\}$ to f, we have that

$$\lim_{i \to \infty} \mu(\{x \in X : |f_i(x) - f(x)| \geq \epsilon') = 0$$

Substituting for ϵ', we recover that

$$\lim_{i \to \infty} \mu(\{x \in X : |f_i(x) - f(x)| \geq \frac{\epsilon}{|c|}) = 0$$

Which further implies that

$$\lim_{i \to \infty} \mu(\{x \in X : |cf_i(x) - cf(x)| \geq \epsilon) = 0$$

Which delivers us the result. ∎

In some sense, one can see these results as indicative that the various modes of convergence act almost like linear operators themselves.

Finally I prove an analogue of the squeeze theorem.

Claim: *Let $\{f_i\}$ is a sequence of $\mathcal{X}$-measurable functions that converge to the zero function along one of the modes of convergence, and let $\{g_i\}$ be a sequence of $\mathcal{X}$-measurable functions. Further suppose that $|g_i| \leq |f_i| \; \forall i$. Then $\{g_i\} \to 0$ along the same mode of convergence.*

Proof. We begin again by considering pointwise convergence. That $\{f_i\} \to 0$ pointwise implies that $\forall \epsilon > 0 \; \forall x \in X \; \exists N_{\epsilon,x}$ such that $\forall i \geq N_{\epsilon,x}$ we have that $|f_i(x) - 0| < \epsilon$. Now observe that since $|g_i(x)| \leq |f_i(x)|$, we also have that $|g_i(x)| - 0 = |g_i(x) - 0| \leq |f_i(x)| - 0 = |f_i(x) - 0|$. Hence $|g_i(x) - 0| < \epsilon$, which delivers the result for pointwise convergence.
As with the prior proof, almost identical arguments will deliver us uniform convergence, μ-almost everywhere convergence of both kinds, and almost uniform convergence.
For convergence in L_1 norm, we observe that $\{f_i\} \to f$ implies that

$$\lim_{i \to \infty} \int_X |f_i(x) - 0| d\mu = 0$$

.
Since we have that $\forall i \; |g_i| \leq |f_i|$ and obviously $0 \leq |g_i|$, by the monotonicity of the unsigned integral, we have that

$$0 = \int_X 0 d\mu \leq \int_X |g_i - 0| d\mu \leq \int_X |f_i - 0| d\mu$$

Taking limits on each side and applying the squeeze theorem[a] delivers the desired result.
The proof for convergence in measure is not entirely dissimilar.
Let $\epsilon > 0$ be given. That $\{f_i\} \to f$ in measure implies that $\forall \epsilon > 0$ we have that $\lim_{i \to \infty} \mu(\{x \in X || f_i(x)| \geq \epsilon\}) = 0$. Since we have $\forall i \; 0 \leq |g_i(x)| \leq |f_i(x)| \forall x \in X$, we obviously have that

$$\forall i \; \{x \in X || g_i(x) \geq \epsilon\} \subseteq \{x \in X || f_i(x)| \geq \epsilon\}$$

By the monotonicity of measure, we have that

$$\forall i \; \mu(\{x \in X || g_i(x) \geq \epsilon\}) \leq \mu(\{x \in X || f_i(x)| \geq \epsilon\})$$

Recalling that measure is by definition non-negative, an application of the squeeze theorem delivers the desired result. ■

[a]The one for sequences that you probably learned in first year calculus.

These four results together give us a pretty potent toolkit for showing that measurable functions converge.[39] But suppose that one wants to further cut the load down when proving convergence. The obvious way to do this is to consider whether or not one kind of convergence implies another.

Claim: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{f_i\}$ be a sequence of $\mathcal{X}$-measurable functions that converge to a $\mathcal{X}$-measurable function f uniformly. Then $\{f_i\}$ converges to f pointwise.*

Proof. The result comes directly from the definition, simply pick $N_{\epsilon,x} = N_\epsilon$. ■

[39]Again, it is important to remember that random variables are simply a special kind of measurable function.

Claim: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{f_i\}$ be a sequence of $\mathcal{X}$-measurable functions that converge to a $\mathcal{X}$-measurable function f uniformly. Then $\{f_i\}$ converges to f in L^∞ norm. Further if $\{f_i\} \to f$ in L^∞, then there exists a null set E such that $\{f_i\} \to f$ uniformly on $X \setminus E$.*

Proof. Both directions are immediate from the respective definitions and the observation that the empty set has measure zero. ∎

Claim: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{f_i\}$ be a sequence of $\mathcal{X}$-measurable functions that converge to a $\mathcal{X}$-measurable function f in L^∞. Then $\{f_i\}$ converges to f almost uniformly.*

Proof. This proof is immediate from the definition. Fix any $\epsilon > 0$. By definition of L^∞ convergence, $\{f_i\} \to f$ uniformly apart from a set of measure zero, so trivially the almost uniform convergence criterion is satisfied. ∎

Claim: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{f_i\}$ be a sequence of $\mathcal{X}$-measurable functions that converge to a $\mathcal{X}$-measurable function f almost uniformly. Then $\{f_i\}$ converges to f pointwise almost everywhere.*

Proof. Fix a sequence of $\nu > 0$ such that ν is monotonically non-increasing and converges to 0 from above. By the almost uniform convergence of the f_i to f, for each ν, the measure of the set where $\{f_i\}$ does not converge to f is at most ν. Call the set associated with a given ν E_i, where $i \in \mathbb{N}$ refers to the index of ν. Invoking downwards monotone convergence delivers us that $\mu(\bigcap_{i=1}^{\infty} = \lim_{i\to\infty} \mu(E_i) = 0$. Now observe that for each E_i, $\{f_i\}$ converges to f uniformly on that set, and thus converges to f pointwise on that set. Since pointwise convergence is preserved under set unions, and the set where pointwise convergence is not guaranteed is measure zero, the result is shown. ∎

Claim: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{f_i\}$ be a sequence of $\mathcal{X}$-measurable functions that converge to a $\mathcal{X}$-measurable function f pointwise. Then $\{f_i\}$ converges to f pointwise almost everywhere.*

Proof. Again the proof is immediate from the definition and the observation that the empty set has measure zero. ∎

Claim: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{f_i\}$ be a sequence of $\mathcal{X}$-measurable functions that converge to a $\mathcal{X}$-measurable function f in L^1 norm. Then $\{f_i\}$ converges to f in measure.*

Proof. Suppose that $\{f_i\} \to f$ in L^1 norm. Let $\epsilon > 0$ be given. By Markov's inequality, we have that

$$\mu\{x \in X || f_i(x) - f(x)| \geq \epsilon\} \leq \frac{1}{\epsilon} \int_X |f_i - f| d\mu$$

By convergence in L^1 norm, we have that $\forall \delta > 0\ \exists N_\delta |\ \forall i \geq N_\delta \int_X |f_i - f| d\mu < \delta$.
Hence we have that

$$0 \leq \mu\{x \in X || f_{N_\delta}(x) - f(x)| \geq \epsilon \leq \frac{1}{\epsilon} \int_X |f_{N_\epsilon} - f| d\mu \leq \frac{\delta}{\epsilon}$$

Taking δ to 0, the result follows. ∎

Claim: *Let $(X, \mathcal{X}, \mu)$ be a measure space, and let $\{f_i\}$ be a sequence of $\mathcal{X}$-measurable functions that converge to a $\mathcal{X}$-measurable function f almost uniformly. Then $\{f_i\}$ converges to f in measure.*

Proof. Suppose that $\{f_i\} \to f$ almost uniformly. Then $\forall \epsilon > 0$, there exists a set E of measure at most ϵ such that such that f_i converges to f uniformly on $X \setminus E$.
Now let $\delta > 0$ be given. Since $\{f_i\} \to f$ almost uniformly, there exists a set E_δ such that the f_i converges to f uniformly on $X \setminus E_\delta$. Now observe that this implies that $\exists N_{\epsilon,\delta}$ such that $\forall i \geq N_{\epsilon,\delta}$ we have that $\forall x \in X \setminus E_\delta\ |f_i(x) - f(x)| < \epsilon$. Hence we have that $\{x \in X || f_i(x) - f(x)| \geq \epsilon\} \subseteq E_\delta$. Invoking the monotonicity of measure, we have that $\mu(\{x \in X || f_i(x) - f(x)| \geq \epsilon\}) \leq \mu(E_\delta) \leq \delta$. Taking δ to 0 delivers the desired result. ∎

It is natural perhaps to think that pointwise almost everywhere convergence implies convergence in measure, but this is not true. As a counterexample, consider the sequence of functions defined by

$$f_i = 1_{[i,i+1]}$$

This sequence converges pointwise to the 0 function, but it does not converge in measure, as for any $i \in \mathbb{N}$, the measure of the set where $f(x) \neq 0$ is exactly equal to 1.

2.8 Stochastic Processes

N.B. The mathematical analysis of stochastic processes (at least as I learned it) lies primarily in the domain of probability theory, and not measure theory writ large. As a consequence, the notation used is that of probability theory, and not of more traditional measure theory as above. That is to say that measurable functions are referred to as random variables, denoted by capital letters, and so on. Hence this is the notation that I make use of for this section of the notes.

Let $\{X_t\}_{t \in T}$ be some collection of random variables from some probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ to some space S (typically for our purposes the reals). The set T, which we require to be totally ordered, is called the index set, and we typically associate it with say, the natural numbers, or the integers, and say that it represents time. The set S is called the **state space**. The entire collection $\{X_t\}_{t\in T}$ is called a **stochastic process**. A sequence of realizations of $\{X_t\}_{t\in T}$ is called a **path**.[40]

Generally, there are three ways of viewing a stochastic process.

1. As a collection of random variables $\{X_t\}$, with one for each time t.
2. As a collection of sample paths $T \to S$, given by $t \to X_t(\omega)$, with one for each ω.
3. As a function from the product space $T \times \Omega \to S$ given by $(t, \omega) \to X_t(\omega)$

But we have a slight problem, in that considering a σ-algebra that is invariant over time places a significant restriction on the information structure that the process can represent. In particular, to the extent that the $\sigma-$algebra $\mathcal{F}$ is invariant for all $t \in T$, we are essentially imposing a restriction the current period information - or the set of possible events - for an agent does not change over time. One way to conceptualize this is that the agent is "small" in a certain sense. To see the restrictiveness of this assumption, suppose that an agent is going to the grocery store. Treat a realization of the stochastic process at time t as an individual's choice of what to buy. By restricting the σ-algebra to be constant, we rule out, among other things, agent's learning about their preferences. To use another, perhaps more frequently seen analogy, suppose we are interested in a stock price. Imposing a constant σ-algebra implies that no new information can be unveiled about the company when valuating the price of the stock. There are no news shocks so to speak.

The way this problem is solved consists of modifying the above probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in T}, \mathbb{P})$. This new collection of objects is called a **Filtered Probability Space** or **Filtered Probability Triple**. A stochastic process is properly defined with respect to this filtered probability space, and not the traditional probability triple. The new object $\{\mathcal{F}_t\}_{t\in T}$ is called a **filtration**: formally, $\{\mathcal{F}_t\}_{t\in T}$ is a sequence of sub-σ-algebras of $\mathcal{F}$ satisfying $\mathcal{F}_s \subseteq \mathcal{F}_t$ $\forall s \leq t$. Generally, when working with a filtration, we would like it to have two properties. The first - **completeness** - is a technical one that says $\mathcal{F}_0$ contains all $\mathbb{P}-$null sets (i.e. all sets of probability zero), and that the underlying triple itself is a complete probability space.[41]. The second - **right continuity** - is the property that $\forall\, t \geq 0$, we have that $\mathcal{F}_t = \mathcal{F}_{t+} = \cap_{s>t}\mathcal{F}_s$. These two conditions are called the **usual conditions**.

How then are we to think of sequences of random variables in this environment? Recall that a random-variable is a function $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{S})$ with the property that $\forall B \in \mathcal{S}$, $X^{-1}(B) \in \mathcal{F}$, or that the pre-image of a measurable set is measurable. Given our motivation of defining information flows over time, the natural way to think about a sequence of random variables is to require that the random variable at time t be measurable with respect to $\mathcal{F}_t$, the sub-σ-algebra at time t. If our process $\{X_t\}$ satisfies this condition for all t, we say that the process is **adapted**. This would mean that the value of $\mathcal{X}_t$ is observable by time t.

To talk about the converse, recall that we say that a random variable X_t generates a σ-algebra on Ω via $\{A \in \Omega | \exists B \in \mathcal{S} \wedge A = X^{-1}(B)\}$. To put it in words, we say that a generated σ-algebra is the $\sigma-$algebra that consists only of the preimages of measurable sets in $(S, \mathcal{S})$. This is by construction a minimal σ-algebra, in the sense that it is the smallest σ-algebra that makes X_t measurable. Naturally then, we can think of about a process as generating a filtration by

[40] Sometimes you will see it referred to as a sample function or a **sample path**.

[41] Recall that the typical example that we give of the Borel σ-algebra on $\mathbb{R}$ is not complete; its completion is the Lebesque σ-algebra

considering the sequence sub σ-algebras generated by the random variables in $\{X_t\}$. This filtration is called the **natural filtration of X**.

If this were a more complete set of notes, then I would put out a number of more interesting results about filtrations, and discuss left and right measurablity, etc.. But they are not, and so I will simply move on to Markov Processes. A Markov Process is a stochastic process that has the **Markov property**, or so called **memorylessness**.

Let $(\Omega, \mathcal{F}, \{F_t\}_{t\geq 0}, \mathbb{P})$ be a filtered probability space, and let $\{X_t\}_{t\geq 0}$ be a stochastic process adapted with respect to $\{\mathcal{F}_t\}$, with $X_t \to (S, \mathcal{S})$. We say that the stochastic process has the Markov property if $\forall B \in \mathcal{S}$ and $s, t \in T$ with $s < t$, the following holds:

$$\mathbb{P}(X_t \in B | X_s) = \mathbb{P}(X_t \in B | \mathcal{F}_s)$$

[a]

[a] In the case where the state space is discrete and the associated σ-algebra is the ordinary one, this reduces to the familiar expression

$$\mathbb{P}(X_t | X_{t-1}, X_{t-2}, \ldots) = \mathbb{P}(X_t | X_{t-1})$$

Formally, the statement above describes a property of first order Markov processes. Informally, the characterization of first order Markov processes is all that matters for tomorrow is what happened today. A second order Markov process then says that all that matters for tomorrow is the state today and the state yesterday. Note that a Markov process of higher order can be transformed into a Markov process of the first order via an appropriate expansion of the state space. Hence, we consider Markov processes of the first order without loss. Note also that the probability measure is not defined with respect to any particular element of the filtration $\{\mathcal{F}_t\}$ but rather with respect to $\mathcal{F}$.

Let us focus for a bit on the discrete time case, as it more familiar and thus somewhat easier to develop intuition in. Let $\{X_t\}$ be a stochastic process in discrete time to some countable state space S with the power σ-algebra. $\{X_t\}$ is a **Markov Chain** if it has the property that $\forall n \in \mathbb{N} \cup \{0\}$ and sequence of states $\{x_i\}_{i=0}^{n+1}$, we have that

$$P(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n)$$

In what should be very unsurprising to you, we often want to make assumptions that make the stochastic process stationary. Formally, we say that the Markov Chain is **stationary** if for all n, k, we have that

$$\mathbb{P}(X_{n+k} = x_{n+k} | X_k = x) = \mathbb{P}(X_n = x_n | X_0 = x)$$

That is to say that the distribution at time $n+k$ conditional on the shock at time k is equivalent to the distribution at time n conditional on the initial shock. In this sense, the initial state is completely irrelevant to the problem, since all that matters is the time between a prior shock and now, and not when that shock actually occurred.

Similarly, we say that a Markov chain is **time homogeneous** if

$$\mathbb{P}(X_{n+1} = x | X_n = y) = \mathbb{P}(X_n = x | X_{n-1} = y)$$

All stationary processes are time homogeneous, which is nice, since stationarity also has the makes it simpler to think about transition matrices. Formally, we define a n-step transition matrix

$$Q_n(x, y) = \mathbb{P}(X_n = y | X_0 = x), \ (x, y) \in S \times S$$

It is easy to see that this is a row stochastic matrix for all $n \in \mathbb{N} \cup \{0\}$. We can also show that $Q_n = Q_1^n$, so we typically are only concerned with the one period transition matrix Q_1, and thus notate it as simply Q.