Introduction to Measure Theoretic Probability

Sam Leone

January 12, 2023

Chapter 1

Motivation & Preliminaries

Why study probability theory? If you're anything like me, you know the basics: dice rolls, taking expectations, and some basic distributions. But along the way, you had a few lingering questions: Why is the law of large numbers true? The central limit theorem? What the heck is a probability density, really? What does it even mean to condition on something with probability 0? There are also the existence of conditioning paradoxes. Perhaps you've heard about stochastic processes and the different limit theorems which show that random processes converge to an equilibrium. How the heck can you show that? We address these through a fundamental shift in perspective: remove the randomness from the study of probability. Rather than studying the "probability" of an event, study its size. The simple idea to use tools from measure theory has far reaching consequences. The added trouble may make you think — why are we doing this? But after a few headaches, the applications will be well worth it.

We will assume a very basic knowledge of real analysis.

1.1 Events, Sizes & The Universe

Let's begin our study of probability with a classic example: the roll of a fair die. We all know that *each side of a die occurs with probability 1/6.* But what do we even mean by this? Some people think of it from a frequentist perspective — if you roll the die 600 billion billion times, it will come up 2 about 100 billion billion times (assuming it doesn't break). In a way, this is silly and circular: we're defining probabilities by the Law of Large Numbers? What could we possibly extend this to densities? If we throw a dart 100 billion billion times, it will most likely hit a given spot 0 times.

The more prudent & careful approach is to think about the universe of possible events, and assign sizes to suitable subsets of those events. Formally, the universe is given by Ω. In the case of the dice roll, you could think of Ω as describing every outcome in every instantiation of this dice roll in the multiverse. The amount of information captured by Ω could be aribtrary - it could contain the outcome of the dice roll, the weather, and what you get for Christmas. But for our game of Monopoly, we don't really care about *everything*. Hence we consider a family of subsets of interest $\mathcal{F}$. For example, $\mathcal{F}$ might consist of the event a 1 is rolled, a 2 is rolled, and all combinations of these. Even if the set of possibilities where a 1 is rolled can be further split up by weather, we essentially turn a blind eye to these distinctions. In our case, $\mathcal{F}$ is called a σ-algebra ("sigma algebra") and has the sensible closure properties.

Definition 1 (σ-algebra)**.** *A family of subsets $\mathcal{F}$ of Ω is called a σ algebra if,*

- $\varnothing \in \mathcal{F}, \Omega \in \mathcal{F}$
- *Closure under complement: For all $A \in \mathcal{F}$, $A^c \in \mathcal{F}$ as well*
- *Closure under countable union: If $\{A_i\}_{i\in I}$ is a countable set such that $A_i \in \mathcal{F}$ for all $i \in I$, then $\cup_{i\in I} A_i \in \mathcal{F}$ as well*

One can verify that these properties imply σ-algebras are also closed under countable union. These requirements are quite natural when considering the operations we normally do in probability.

Finally, we need a machine which computes sizes. This is done through a so-called *measure* μ. So in the dice roll, $\{\text{roll a } 6\} \in \mathcal{F}$, and $\mu(\{\text{roll a } 6\}) = 1/6$. Formally, μ can be regarded as a set $\mu : \mathcal{F} \to \mathbb{R}^+$. Note that, in the probabilistic case, $\mu(\Omega) = 1$ (tee size of everything is 1), but this need not be true in general. In fact, we will consider a notable exception: the Lebesgue measure. The necessary properties of μ pair nicely with the definition of a σ-algebra. In short, we require nonnegative sizes, the size of nothing to be 0, and that the sizes of non-overlapping things adds.

Definition 2 (Measure)**.** *A (countably-additive) measure on $\mathcal{F}$ is a function $\mu : \mathcal{F} \to \mathbb{R}$ such that,*

- *For all $A \in \mathcal{F}, \mu(A) \geqslant 0$*
- $\mu(\varnothing) = \varnothing$
- *If $\{A_i\}_i$ are countable in $\mathcal{F}$ and pairwise disjoint, then $\mu(\cup_{i\in I} A_i) = \sum_{i\in I} \mu(A_i)$*

In particular, if $\mu(\Omega) = 1$, μ is a *probability measure*. Hopefully, the first two conditions are clear and well-motivated. For the last one, we are simply requiring something like $\mu(\{\text{roll a } 1 \cup \{\text{roll a } 2\}\}) = \mu(\{\text{roll a } 1\}) + \mu(\{\text{roll a } 2\})$. There are a few special cases worth familiarizing ourselves with: probability measures $\subseteq$ finite measures $\subseteq$ σ-finite measures:

Definition 3 (Finite Measure)**.** *If $\mu(\Omega) < \infty$, then μ is finite.*

Definition 4 (Probability Measure)**.** *If $\mu(\Omega) < \infty$, then μ is a probability measure.*

Definition 5 (σ-finite)**.** *If $\Omega = \cup_{i \in I} A_i$, such that I is countable and each $A_i \in \mathcal{F}$ but $\mu(A_i) < \infty$, then μ is σ-finite*

Definition 6 (Measure Space)**.** *A measure space is a triple $(\Omega, \mathcal{F}, \mu)$ where $\mathcal{F}$ is a σ-algebra of subsets of Ω, and μ is a measure on $\mathcal{F}$*

1.2 σ-algebras and Generating Sets

Suppose $\mathcal{E}$ is a family of subsets of Ω (now, we make no assumptions on the nature of $\mathcal{E}$). We say the σ-algebra generated by $\mathcal{E}$, denoted $\sigma(\mathcal{E})$ is the smallest σ-algebra containing $\mathcal{E}$:

$$\sigma(\mathcal{E}) = \bigcap_{\sigma\text{-algebras } \mathcal{F} s.t. \mathcal{E} \subseteq \mathcal{F}} \mathcal{F}$$

In this sense, $\mathcal{E}$ can be thought of as the atoms of Ω from which we build molecules in $\mathcal{F}$. As we will see, these atoms need not be unique. For example, if $\Omega = \{1, 2, 3, 4, 5, 6\}$, if $\mathcal{E} = \{\{1\}, \{2\}, \{3\}...\{6\}\}$, then $\sigma(\mathcal{E}) = \mathcal{P}(\Omega)$, the full power set. However, if $\mathcal{E} = \{\{1, 2, 3\}, \{4, 5, 6\}\}$, then the resulting structure has a "lower resolution:" $\sigma(\mathcal{E}) = \{\varnothing, \{1, 2, 3\}, \{4, 5, 6\}, \Omega\}$. Note that the following intuitive properties hold:

Lemma 1.2.1. *Let $\mathcal{E} \subseteq \mathcal{P}(\Omega)$. Then,*

- *If $\mathcal{E}_1$ is a σ-algebra, then $\sigma(\mathcal{E}) = \mathcal{E}$*
- *If $\mathcal{E} \subseteq \mathcal{E}'$, then $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E}')$*

Proof. Clearly, as $\mathcal{E}$ is a σ-algebra containing $\mathcal{E}$,

$$\sigma(\mathcal{E}) = \mathcal{E} \cap \bigcap_{\sigma\text{-algebras } \mathcal{F} s.t. \mathcal{E} \subseteq \mathcal{F}} \mathcal{F} \subseteq \mathcal{E}$$

Also, $\sigma(\mathcal{E}) \supseteq \mathcal{E}$ by definition. We conclude $\sigma(\mathcal{E}) = \mathcal{E}$. For the latter claim, we prove $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E}')$ by considering arbitrary elements of $\sigma(\mathcal{E})$. Suppose $A \in \sigma(\mathcal{E})$. By definition, for all $\mathcal{F}$ containing $\mathcal{E}$, $A \in \mathcal{F}$. Note also that for all $\mathcal{F}'$ containing $\mathcal{E}'$, $\mathcal{F}'$ contains $\mathcal{E}$ as well, so $A \in \mathcal{F}'$. As a consequence $A \in \sigma(\mathcal{E}')$. Since A was generic, $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E}')$.

□

1.2.1 The Borel σ-algebra

It would be no overstatement to say the most often studied σ-algebra is the Borel σ-algebra. In this case, $\Omega = \mathbb{R}$, $\mathcal{E}$ = the open sets in $\mathbb{R}$, and $\mathcal{F} = \sigma(\mathcal{E})$. This is denoted $\mathcal{B}(\mathbb{R})$. More generally, the Borel σ-algebra of a metric space $\mathcal{X}$ is denoted $\mathcal{B}(\mathcal{X})$. Recall from definition 1 that $\mathcal{B}(\mathbb{R})$ should be closed under compliment, and so $\mathcal{B}(\mathbb{R})$ contains the closed sets as well. Moving forward, the Borel σ-algebra will contain all the richness we will practically need.

1.2.2 Lebesgue-Stieltjes Measures

A generic class of measures is the set of Riemann-Stieltjes Measures. A distribution function (think CDF) is a map $F : \mathbb{R} \to \mathbb{R}$ such that,

- $F(x)$ is nondecreasing with x
- F is right continuous ($\lim_{y \to x^+} F(y) = F(x)$)

The corresponding Lebesgue-Stieltjes measure sets $\mu((a, b]) = F(b) - F(a)$. It can be shown that this is enough to specify the whole measure over $\mathcal{B}(\mathbb{R})$.

Example 1 (The Lebesgue Measure)**.** *The Lebesgue measure is the Lebesgue-Stieltjes Measure when $F(x) = x$, and so $\mu((a, b]) = b - a$.*

Example 2 (The Normal Distribution)**.** *The normal distribution is induced by the Lebesgue-Stieltjes measure with $\mu((a, b]) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$.*

Example 3 (Probability Mass Functions from CDFs)**.** *More generally, when F is the CDF of a probability mass function, the corresponding Lebesgue-Stieltjes measure corresponds to that probability distribution.*

1.3 All We Care About - Negligible Sets & Almost Everywhereness

To proceed with our study of measure theory, and thus probability theory, we will make use of the notion of something happening almost everywhere. We say an event $A \in \mathcal{F}$ occurs almost everywhere w.r.t a measure μ if $\mu(A^c) = 0$. Likewise, N is said to be a negligible set if $\mu(N) = 0$.

Proposition 1 (Properties of Negligible Sets)**.** *Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space.*

- *If A is negligible and $B \subseteq A$, then B is a negligible set*
- *If $A_1, A_2 ...$ are negligible sets, then so is $\cup_{i=1}^{\infty} A_i$*

Chapter 2

Extending Lebesgue-Stieltjes Measures & Carathedory's Theorem

Note that if you know you have a measure, you can do everything your heart desires! This chapter it not dedicated to the existence of measures so much as proving that we can construct measures with desireable properties.

2.1 The Problem

You are a contractor. Your client comes to you, embarassed, and says *hey, I have my Lebesgue-Stieltjes measure* μ*. Could you help me define it uniquely over all of* $\mathcal{B}(\mathbb{R})$? You say *sure thing, let me just plug it into my measure extender machine*. And out pops a new measure consistent, defined over new sets, and it is consistent with the old one. The main example to bear in mind is the Lebesgue measure.

Begin by saying that you have a measure μ_S (S for Start) defined on an *semi-ring* $\mathcal{A}$. While we have not yet defined semi-ring, think of it as an incredibly simple family of sets. For example, intervals of the form $(a, b]$ comprise a semi-ring; we could specify a Lebesgue-Stieltjes function on this semi-ring. d-dimensional boxes like $\times_{i=1}^{d}(a_i, b_i]$ also comprise a semi-ring. The problem is basically this: your customer comes to you and says *hey buddy, I already know what I want* μ_S *to be like on* $\mathcal{A}$*, can you help me out on* $\sigma(\mathcal{A})$*?* You say, probably! More formally, we seek to prove a theorem roughly of the form:

Theorem 2.1.1. *Under assumptions, given a baby measure* μ_S *defined on a semi-ring* $\mathcal{A}$*, there exists a unique measure* μ *defined on* $\sigma(\mathcal{A})$ *which respects* μ_S*.*

2.2 Semi-Rings & Rings

This will be boring, but necessary. We are going to define a series of related families of sets. Let Ω be the universe and $\mathcal{A}$ be a family of subsets of Ω. The relationships to bear in mind are:

$$\text{Semi-Ring} \implies \text{Ring} \implies \text{Algebra} \implies \sigma - \text{Algebra}$$

Definition 7 (Semi-Ring)**.** *A family of sets $\mathcal{A}$ is said to be a semi-ring if, for all $A, B \in \mathcal{A}$,*

- $\varnothing \in \mathcal{A}$
- $A \cap B \in \mathcal{A}$
- $A \backslash B = \cup_{1 \leqslant j \leqslant n} C_j$*, where* $C_j \in \mathcal{A}$ *and the* C_j*'s are all pairwise disjoint.*

Proposition 2. *$\mathcal{I}$ is a semi-ring.*

The canonical example to bear in mind is the semi-ring of half-open intervals. Define $\mathcal{I}$ to be all the sets of the form $(a, b]$ with $a \leqslant b$ in $\mathbb{R}$. So $\mathcal{I} = \{(a, b] : a, b \in \mathbb{R}\}$.

Proof. Letting $a = b$, $(a, b] = \varnothing$. Properties (ii) and (iii) can be checked by simple casework on any two intervals $A = (a, b], B = (c, d]$. ☐

Note that semi-rings, in particular this semi-ring, is not closed under union. If we add this property, we obtain a ring.

Definition 8 (Ring)**.** *A family of sets $\mathcal{A}$ is said to be a ring if, for all $A, B \in \mathcal{A}$,*

- $\varnothing \in \mathcal{A}$
- $A \cup B$
- $A \backslash B \in \mathcal{A}$

Just like how we defined the σ-algebra generated by a set, the ring generated by $\mathcal{A}$ is the smallest ring containing $\mathcal{A}$.

Proposition 3. *Let $\mathcal{A}$ be a semi-ring. Let $\mathcal{B}$ be the set of finite disjoint unions of elements of $\mathcal{A}$. Then $\mathcal{B} = ring(\mathcal{A})$*

Proof. We begin by showing $\mathcal{B}$ is a ring. Note that $\varnothing \in \mathcal{B}$ as $\varnothing \in \mathcal{A}$, so it can be considered as the union of one element of $\mathcal{A}$. Now, write $A = \bigcup_{i=1}^n A_i, B = \bigcup_{j=1}^m B_j$. We shall show that $A \cup B \in \mathcal{A}$. Indeed, $A \cup B = \bigcup_{i=1}^n A_i \cup \bigcup_{j=1}^m B_j$, which is also an element of $\mathcal{A}$ by definition. Now, it remains to show that $A \backslash B \in \mathcal{A}$. To see this, note that,

$$A \backslash B = \bigcup_{i=1}^{n} A_i \backslash \bigcup_{j=1}^{m} B_j$$

This can be understood as those elements x which belong to at least one A_i, but not a single B_j. From this, it's clear that this can be understood as:

$$\bigcup_{i=1}^{n} \bigcap_{j=1}^{m} A_i \backslash B_j$$

Note also that by definition of a semi-ring, we can write $A_i \backslash B_j = \cup_{k=1}^{n_{i,j}} C_{i,j,k}$, where these are all disjoint. And thus, we have,

$$A \backslash B = \bigcup_{i=1}^{n} \bigcap_{j=1}^{m} \bigcup_{k=1}^{n_{i,j}} C_{i,j,k}$$

Now there's a bit of a subtle thing going on. As the $C_{i,j,k}$'s are pairwise disjoint for fixed k, if x is in $\bigcap_{j=1}^{m} \bigcup_{k=1}^{n_{i,j}} C_{i,j,k}$, it belongs to precisely one $C_{i,j,k}$ for each j. Let $\mathcal{C}_{i,j} = \{C_{i,j,k} : 1 \leqslant k \leqslant n_{i,j}\}$. We may then write:

$$\bigcap_{j=1}^{m} \bigcup_{k=1}^{n_{i,j}} C_{i,j,k} = \underbrace{\bigcup_{C_1 \in \mathcal{C}_{i,1} \ldots C_{i,m} \in \mathcal{C}_{i,m}} C_1 \cap C_2 \ldots \cap C_m}_{\in \mathcal{B}}$$

Since $\mathcal{A}$ is a semi-ring and thus is closed under intersection, each $C_1 \cap C_2 .. \cap C_m$ is in $\mathcal{A}$, so the union of such intersections is in $\mathcal{B}$. Thus, $\bigcap_{j=1}^{m} \bigcup_{k=1}^{n_{i,j}} C_{i,j,k}$ is in $\mathcal{A}$ for all i. And so, as we've already shown $\mathcal{B}$ is closed under union, $A \backslash B \in \mathcal{B}$. This proves the desired property. And so $\mathcal{B}$ is indeed a ring.

To see that $\mathcal{B} = \text{ring}(\mathcal{A})$ is simple. As $\text{ring}(\mathcal{A})$ is a ring, it must contain all unions of elements of $\mathcal{A}$, so $A \subseteq \mathcal{B} \subseteq \text{ring}(\mathcal{A})$. Taking the ring of all sides, and noting $\text{ring}(\mathcal{B}) = \mathcal{B}$, as $\mathcal{B}$ is a ring, $\text{ring}(\mathcal{A}) \subseteq \mathcal{B} \subseteq \text{ring}(\mathcal{A})$, so $\mathcal{B} = \text{ring}(\mathcal{A})$.

□

2.3 Extending Lebesgue-Stieltjes Measures from Semi-Rings to Rings

We now show that we can extend measures from semi-rings to rings. First, we define a relaxed version of a measure that we care about.

Definition 9 (finitely additive measure)**.** *Suppose $\mu : \mathcal{A} \to \mathbb{R}$ is a function on subsets of Ω. We say μ is finitely additive if,*

- $\mu(\varnothing) = 0$
- *If $A \subseteq B, \mu(A) \leqslant \mu(B)$*
- *If $A, B \in \mathcal{A}$, $A \cap B = \varnothing$ and $A \cup B \in \mathcal{A}$, then $\mu(A \cup B) = \mu(A) + \mu(B)$*

Theorem 2.3.1. *Let μ_S be a countably additive measure defined on a semi-ring $\mathcal{A}$. Let $\mathcal{B} = ring(\mathcal{A})$. Then there exists a unique finitely additive measure μ acting on $\mathcal{B}$ which respects μ_S over $\mathcal{A}$.*

Proof. We provide an explicit construction for μ, then verify that all is well. For an arbitrary element $B \in \mathcal{B}$, let $B = \cup_{i=1}^{n} A_i$ (we know from Proposition 3 that this is the form of such elements). And assume without loss of generality that the A_i's are disjoint. Why can we do this? Note for $A, B \in \mathcal{A}$, we have $A \cup B = B \cup (A \backslash B) = B \cup \cup_{i=1}^{m} C_i$, where the C_i's are pairwise disjoint. By induction, then, every finite union can be represented as a finite disjoint union. We do the perfectly natural thing: we want our measure to be finitely additive, so our hand is forced. We define,

$$\mu(B) = \sum_{i=1}^{n} \mu_S(A_i)$$

First, we must verify that this is consistent. Suppose $B = \cup_{j=1}^{m} B_i$. Then observe,

$$\sum_{i=1}^{n} \mu_S(A_i) = \sum_{i=1}^{n} \mu_S(A_i \cap B) = \sum_{i=1}^{n} \mu_S(A_i \cap \bigcup_{j=1}^{m} B_j)$$
$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \mu_S(A_i \cap B_j) = \sum_{j=1}^{m} \mu_S(B_j)$$

Where we have employed the assumption that μ_S is finitely additive over $\mathcal{A}$ and utilized the fact that semi-rings are closed under intersection (and so $\mu_S(A_i \cap B_j)$ is defined). From here, it is trivial to verify that μ is finitely additive. Indeed, we simply seek to show that if $A, B \in \mathcal{B}$ are disjoint, then $\mu(A \cup B) = \mu(A) \cup \mu(B)$. First, write $A = \cup_{i=1}^{n} A_i, B = \cup_{j=1}^{m} B_j$. Again, assume the A_i's are pairwise disjoint, as are the B_j's. But since $A \cap B = \varnothing$, the A_i's are also disjoint with the B_j's. So then,

$$\mu(A \cup B) = \mu\left(\bigcup_{i=1}^{n} A_i \cup \bigcup_{j=1}^{m} B_j\right) = \sum_{i=1}^{n} \mu(A_i) + \sum_{j=1}^{m} \mu(B_j) = \mu(A) + \mu(B)$$

Which proves the additivity property. By induction, one can easily establish that if $B_1, B_2 ... B_n \in \mathcal{B}$ are all pairwise disjoint, then $\mu(\cup_{i=1}^{n} B_i) = \sum_{i=1}^{n} \mu(B_i)$.

Uniqueness of μ is trivial. If there is a second μ' that respects μ_S, then countable additivity forces $\mu = \mu'$ over everything in $\mathcal{B}$.

□

At this point, we would like to show that μ is not only finitely additive, but countably additive over $\mathcal{B}$. The following proposition ensures that countable additivity of μ_S over $\mathcal{A}$ is sufficient.

Theorem 2.3.2. *Letting $\mathcal{A}$ be a semi-ring and let $\mathcal{J}$ be the ring generated by $\mathcal{A}$. Let μ_S, μ as described in the above theorem. Then if μ is countably additive over $\mathcal{I}$, then μ is countably additive over $\mathcal{J}$.*

Proof. First, as $B \in \mathcal{B}$, we may write $B = \cup_{j=1}^m A_j$, where the A_j's are pairwise disjoint. So let us first restrict our analysis to a particular A_j. Note that $B = \cup_{i=1}^\infty B_i$ as well. Note that, as the A_j's are pairwise disjoint, as are the B_i's, it must be the case that each A_j is a collection of the B_i's. So let I_j be such that $A_j = \cup_{i \in I_j} B_i$. Note that, if we can prove that $\mu(A_j) = \sum_{i \in I_j} \mu(B_i)$ for each j, we will be done, since then finite additivity will imply,

$$\mu(B) = \sum_{j=1}^m \mu(A_j) = \sum_{j=1}^m \sum_{i \in I_j} \mu(B_i) = \sum_{i=1}^\infty \mu(B_i)$$

So it remains to prove $\mu(A_j) = \sum_{i \in I_j} \mu(B_i)$ for arbitrary A_j. Finally, observe that B_i can be written as $\cup_{k=1}^{n_i} C_{i,k}$. And so, $A_j = \cup_{i \in I_j} \cup_{k=1}^{n_i} C_{i,k}$. Then by countable additivity of μ_S on $\mathcal{I}$,

$$\sum_{i \in I_j} \mu(B_i) = \sum_{i \in I_j} \sum_{k=1}^{n_i} \mu_S(C_{i,k}) = \mu(A_j)$$

□

Now, at this point, it will prove somewhat difficult to prove theorems in full generality. So let us abandon our hope of working with completely arbitrary semi-rings. From now on, we will let $\mathcal{I}$ be the semi-ring of d-dimensional boxes $\{(a_1, b_1] \times (a_2, b_2] ... \times (a_d, b_d] : a_1, b_1 ... a_d, b_d \in \mathbb{R}\}$.

Proposition 4. *$\mathcal{I}$ as described is a semi-ring*

Proof. Exercise □

It is worth asking: when is μ_S actually countably additive? Not any set function will do. For instance, if $\mu_S(a, b] = 2^{b-a}$, even though this satisfies the monotonicity property and $\mu_S(\varnothing) = 0$, additivity crumbles. Thus, we will restrict our study to d-dimensional Lebesgue-Stieltjes measures. That is, we assume the existence of d distribution functions $F_1 ... F_d$, and set,

$$\mu_S\left(\bigtimes_{i=1}^d (a_i, b_i] \right) = \prod_{i=1}^d (F_i(b_i) - F_i(a_i))$$

Theorem 2.3.3. *μ as described is countably additive over $\mathcal{I}$ and thus its extension to $\mathcal{J}$ is countably additive as well.*

Proof. Assume $A = (a_1, b_1] \times ... \times (a_d, b_d] \in \mathcal{I}$. Also assume we have $A = \bigcup_{i=1}^\infty A_i$, where each $A_i = \bigtimes_{j=1}^d (a_{i,j}, b_{i,j}]$. We seek to show,

$$\mu_S(A) = \sum_{i=1}^n \mu_S(A_i)$$

We will argue this via induction on d.

Base Case: First, suppose $d = 1$, so we consider a measure on the real line. First, observe μ_S is then finitely additive. If $(a, b]$ and $(c, d]$ are disjoint (assume $a < c$), then for $(a, b] \cup (c, d] \in \mathcal{I}$, we have $b = c$, so $\mu_S((a, b] \cup (c, d]) = \mu_S((a, d]) = F_1(d) - F_1(a)$. Likewise, $\mu_S((a, b]) + \mu_S((c, d]) = F_1(d) - F_1(c) + F_1(b) - F_1(a) = F_1(d) - F_1(a)$. We use this finite additivity over and over again. Note indeed, that

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

Taking $n \to \infty$, we find $\sum_{i=1}^{\infty} \mu(A_i) \leqslant \mu(A)$. We now seek to show the reverse inequality. Fix any $\epsilon > 0$ and consider an augmentation of the A_i's. Let $\delta > 0$ and $\delta_1, \delta_2.... > 0$ be arbitrary for now. $A_i = (a_i, b_i]$, let $A_i' = (a_i, b_i + \delta_i]$. Also consider, where $A = (a, b]$, consider the new interval $A' = [a + \delta, b]$. As $\{A_i'\}_i$ provides a covering of A, it provides a covering of A'. Thus, it is possible to extract a finite cover. So let $I \subseteq \mathbb{N}$ be such that $A' \subseteq \cup_{i \in I} A_i'$.

Proposition 5. *If $A, B \in \mathcal{B}$, then $\mu(A \cup B) \leqslant \mu(A) + \mu(B)$*

Proof. By additivity,

$$\mu(A \cup B) = \mu(A) + \mu(B \backslash A) \leqslant \mu(A) + \mu(B)$$

By induction, this holds for any finite number of sets as well. □

We can use this fact to upper bound $\mu(A')$ by a nondisjoint union:

$$\mu(A') \leqslant \mu\left(\bigcup_{i \in I} A_i'\right) \leqslant \sum_{i \in I} \mu(A_i')$$

$$= \sum_{i \in I} F(b_i + \delta_i) - F(a) = \sum_{i \in I} (F(b_i + \delta_i) - F(b_i) + F(b_i) - F(a))$$

$$\leqslant \sum_{i=1}^{\infty} (F(b_i) - F(a_i)) + (F(b_i + \delta_i) - F(b_i))$$

$$= \sum_{i=1}^{\infty} \mu(A_i) + \sum_{i=1}^{\infty} (F(b_i + \delta_i) - F(b_i))$$

Additionally,

$$\mu(A) = F(b) - F(a) = F(b) - F(a + \delta) + F(a + \delta) - F(a) = \mu(A') + F(a + \delta) - F(a)$$

By right continuity of F, we can let δ be such that $F(a + \delta) - F(a) < \epsilon/2$. We may also let each δ_i be such that $F(b + \delta_i) - F(b) < \epsilon/2^i$. So then,

$$\mu(A) < \mu(A') - \epsilon/2 \leqslant \sum_{i=1}^{\infty} \mu(A_i) + \sum_{i=1}^{\infty} \epsilon/2^i - \epsilon/2$$

$$= \sum_{i=1}^{\infty} \mu(A_i) + \epsilon/2$$

At last, taking $\epsilon \to 0$, we find $\mu(A) \leqslant \sum_{i=1}^{\infty} \mu(A_i)$. This completes the proof of the base case.

Inductive Step: Now, we assume that this theorem is true in $d-1$ dimensions, and we seek to push it to d dimensions. So suppose that $A = \cup_{i=1}^{\infty} A_i$, where the A_i's are disjoint. Now, without loss of generality, suppose that we take a common refinemenet of the A_i's in the dth dimension. That is, let $H = \bigcup_{i=1}^{n}(a_{i,d} \cup b_{i,d})$ be the set of all numbers which are relevant to our partition along dimension d. Now put H in increasing order, such that: $H = \{c_1 < c_2 ... < c_n\}$. So now, split the A_i's by H into a new collection A_1', A_2'.... Each A_i' can be written as $A_i' = \times_{i=1}^{d-1}(a_i', b_i'] \times (c_k, c_{k+1}]$ for some k. Now, partition the A_i''s into sets I_1, I_2... such that for $i \in I_k$, A_i's dimension d component looks like $(c_k, c_{k+1}]$. So then,

$$A = \cup_{k=1}^{\infty} \cup_{i \in I_k} A_i'$$

Define a new distribution function G where $G(c_k) = \sum_{1 \leqslant j \leqslant k} \mu(\cup_{i \in I_k} A_i')$. G thus corresonds to a 1-dimensional distribution function. So by the base case, we have,

$$\mu(A) = \sum_{k=1}^{\infty} \mu\left(\bigcup_{i \in I_k} A_i' \right)$$

Now, let μ_{d-1} be the measure induced by considering the first $d-1$ dimensions of the A_i's. We have $\mu\left(\bigcup_{i \in I_k} A_i' \right) = (c_{k+1} - c_k)\mu_{d-1}\left(\bigcup_{i \in I_k} A_i' \right)$. So then, by induction,

$$\sum_{k=1}^{\infty} \mu\left(\bigcup_{i \in I_k} A_i' \right) = \sum_{k=1}^{\infty} (c_{k+1} - c_k)\mu_{d-1}\left(\bigcup_{i \in I_k} A_i' \right)$$

$$= \sum_{k=1}^{\infty} (c_{k+1} - c_k) \sum_{i \in I_k} \prod_{i=1}^{d-1} (a_i, b_i] = \sum_{k=1}^{\infty} \sum_{i \in I_k} \prod_{i=1}^{d} (a_i, b_i] = \sum_{i=1}^{\infty} \mu(A_i')$$

Note that the same decomposition into the dth and first $d-1$ dimensions yields:

$$\mu(A_i) = \sum_{i : A_i' \subseteq A_i} \mu(A_i')$$

Which collectively implies that $\mu(A) = \sum_{i=1}^{\infty} \mu(A_i)$. This completes the inductive step and thus the whole proof.

$\square$

Corollary 2.3.3.1. *If μ_S is a d-dimensional Lebesgue-Stieltjes measure on $\mathcal{I}$, then μ is countably additive on $\mathcal{J}$*

2.3.1 Recap

Let's pause for a moment to focus on what we've actually done. We have shown that if we have a d-dimensional distribution function, we can extend it to a countably additive measure on a ring. We will find that rings are very nice. In particular, rings can approximate sets in the Borel σ algebra arbitrarily well. This is the fact we will use to define an outer measure.

2.4 Outer Measures

Thus far, we have worked our way "up" from our semi-ring and tried to build up something more sophisticated on rings, a more complicated family of sets. Now, we will develop a sort of master function, called a *outer measure*, which is indeed defined on all subsets of Ω and thus $\mathcal{B}(\mathbb{R}^d)$ as well. While outer measures do not behave well in general, we will show that it acts nicely on $\mathcal{B}(\mathbb{R}^d)$.

Recall we have a measure μ acting on the ring $\mathcal{J}$ generated by the half-open intervals. We define the following outer measure on subsets of $\mathbb{R}^d$:

$$\mu^\star(A) = \inf\{\mu(J) : J \in \mathcal{J}, A \subseteq J\}$$

Intuitively, the idea is this: we wrap an element J of $\mathcal{J}$ around A as tightly as possible, and then take sizes the way we know how. Then like shrink wrap, we make J as small as possible. First, let us establish a key fact.

Proposition 6. *If $A \in \mathcal{J}$, then $\mu^\star(A) = \mu(A)$*

Proof. Obviously, as A is a valid candidate from $\mathcal{J}$, $\mu^\star(A) \leqslant \mu(A)$. Now we show the reverse. Suppose by way of contradiction that $\mu^\star(A) < \mu(A)$. Then there would exist a $B \in \mathcal{J}$ with $A \subseteq B$ such that $\mu(B) < \mu(A)$. This is of course a contradiction of the monotonicity property. $\square$

Now, we define a very general family of sets: the Lebesgue measurable sets. This will turn out to be more general than we need.

Definition 10 (Lebesgue Measurable)**.** *Say a set $E \subseteq \mathbb{R}$ is Lebesgue-measurable if it satisfies the Caratheodory criterion: that for all $A \subseteq \mathbb{R}$, $\mu^\star(A) = \mu^\star(A \cap E) + \mu^\star(A \cap E^c)$. Let the Lebesgue measurable sets be $\mathcal{L}$.*

As a first observation, note that proposition 6 and Corollary 2.3.3.1 collectively imply that $\mu^\star$ is countably additive on $\mathcal{J}$. Here are the remaining steps:

1. Prove that $\sigma(\mathcal{J}) = \mathcal{B}(\mathbb{R}^d)$
2. Observe $\mathcal{J} \subseteq \mathcal{L}$
3. Prove $\mu^\star$ is countably additive on $\mathcal{L}$

4. Prove $\mathcal{L}$ is a σ-algebra
5. Deduce $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{J}) \subseteq \sigma(\mathcal{L}) = \mathcal{L}$

From this, it will follow that $\mu^\star$ is a countably additive measure on $\mathcal{B}(\mathbb{R}^d)$, so we will be done.

2.4.1 Step 1

Theorem 2.4.1. $\sigma(\mathcal{J}) = \mathcal{B}(\mathbb{R}^d)$

Proof. Let us first prove $\mathcal{I} \subseteq \mathcal{B}(\mathbb{R})$ for $d = 1$. Indeed, note,

$$(a, b] = \cap_{n=1}^{\infty}(a, b + 1/n)$$

And so each $(a, b] \in \mathcal{B}(\mathbb{R})$. Thus, $\mathcal{I} \subseteq \mathcal{B}(\mathbb{R})$. It remains to show the reverse inclusion.

Proposition 7. *Every open set in $\mathbb{R}$ can be written as the disjoint union of open intervals*

Proof. Let O be open. Let $O_Q = O \cap \mathbb{Q}$. Observe that by definition of opennes, for each $q \in O_Q$, there exists a highest $\epsilon_q > 0$ such that $B_{\epsilon_q}(q) \subseteq O$. Now, let $C = \{B_{\epsilon_q}(q) : q \in O_Q\}$ and $S = \cup_{I \in C} I = O$. I claim $O = S$. To see this, observe for any $x \in O$ that there is an ϵ ball $B_\epsilon(x)$ contained in O. If we let q be a rational number s.t. $|x - q| < \epsilon/2$, by maximality of ϵ_q, we have $x \in B_{\epsilon_q}(q)$, so $x \in C$. Finally, let elements of D be obtained by connecting all intervals in C, so that D consists of disjoint open intervals and $\cup_{I \in D} I = \cup_{I \in C} I = S = O$. Thus, O can be written as the disjoint open intervals provided in D. □

Thus, letting $O \in \mathcal{G}$ be some arbitrary open set in $\mathbb{R}$, where we know $\mathcal{G}$ generates $\mathcal{B}(\mathbb{R}^d)$. Thus, the set of open intervals, call it $\mathcal{E}$, generates $\mathcal{B}(\mathbb{R})$. Yet also, any interval (a, b) can be written as:

$$(a, b) = \cup_{n=1}^{\infty}(a, b - 1/n]$$

Which is in $\sigma(\mathcal{I})$. Thus, $\mathcal{E} \subseteq \sigma(\mathcal{I})$, so $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{I})$. This is sufficient to prove the claim in dimension $d = 1$. It remains to prove it for higher dimensions.

Proposition 8. *Let $\mathcal{E}$ generate $\mathcal{F}$. Then*

$$\sigma(\{A_1 \times A_2 ... \times A_n : A_1 ... A_n \in \mathcal{E}\}) = \sigma(\{A_1 \times A_2 ... \times A_n : A_1 ... A_n \in \mathcal{F}\})$$

Proof. Exercise. It is best to prove this when $d = 2$ and proceed by induction. □

Corollary 2.4.1.1. *If $\mathcal{E}_1, \mathcal{E}_2$ both generate $\mathcal{F}$, then,*

$$\sigma(\{A_1 \times A_2 ... \times A_n : A_1 ... A_n \in \mathcal{E}_1\}) = \sigma(\{A_1 \times A_2 ... \times A_n : A_1 ... A_n \in \mathcal{F}\})$$
$$= \sigma(\{A_1 \times A_2 ... \times A_n : A_1 ... A_n \in \mathcal{E}_2\})$$

Proposition 9. *Every open set in $\mathbb{R}^d$ can be written as the (not necessaarily disjoint) union of countably many open rectangles*

Proof. Following the same outline as before, use the fact that $\mathbb{Q}^d$ is dense in $\mathbb{R}^d$ and the definition of the open sets. The reason we no longer have disjointness is that the union of two connected open rectangles may not be an open rectangle in dimension greater than 1. $\square$

A corollary of this is that $\mathcal{B}(\mathbb{R}^d)$ is generated by the set of open rectangles.

Corollary 2.4.1.2. $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R}^d)$ *for all* d.

Proof. Let $\mathcal{E}_1 = \{(a, b] : a, b \in \mathbb{R}\}$, which is simply $\mathcal{I}$ in dimension 1. Also let $\mathcal{E}_2 = \{(a, b) : a, b \in \mathbb{R}\}$. Then,

$$\sigma(\mathcal{I}) = \sigma(\{I_1 \times I_2 .. \times I_d : I_1, ..I_d \in \mathcal{E}_1\})$$
$$= \sigma(\{A_1 \times A_2 ... \times A_d : A_1 .. A_d \in \sigma(\mathcal{E}_2)\}) = \mathcal{B}(\mathbb{R}^d)$$

$\square$

And of course, $\sigma(\mathcal{I}) = \sigma(\mathcal{J})$. This concludes the proof.

$\square$

2.4.2 Step 2

Theorem 2.4.2. $\mathcal{J} \subseteq \mathcal{L}$

Proof. Consider arbitrary $E \in \mathcal{J}$. Now consider arbitrary $A \subseteq \mathbb{R}^d$. We desire to show that

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

We will prove this by showing the two corresponding inequalities. Let J_1, J_2 be such that,

$$\mu(J_1) < \mu^\star(A \cap E) + \epsilon/2$$
$$\mu(J_2) < \mu^\star(A \cap E^c) + \epsilon/2$$

Let $J = J_1 \cup J_2$. It's clear that $J \in \mathcal{J}$. Furthermore, as $A \cap E \subseteq J, A \cap E^c \subseteq J, A = (A \cap E) \cup (A \cap E^c) \subseteq J$. So then we have,

$$\mu^\star(A) \leqslant \mu^\star(J) \leqslant \mu(J_1) + \mu(J_2) < \mu^\star(A \cap E) + \mu^\star(A \cap E^c) + \epsilon$$

Taking $\epsilon \to 0$, we have side of the equality. Now, we show the reverse inequality. We seek to show,

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

Suppose that $A \subseteq J$. Then $A \cap E \subseteq J \cap E$. Furthermore, $A = (A \cap E) \cup (A \cap E^c) \subseteq (J \cap E) \cup (J \cap E^c)$. And thus,

$$\mu^\star(A) + \epsilon \geqslant \mu(J) = \mu(J \cap E) + \mu(J \cap E^c)$$

But note that $J \cap E, J \cap E^c \in \mathcal{J}$, so,

$$\geqslant \mu(A \cap E) + \mu(A \cap E^c)$$

Taking $\epsilon \to 0$, we're done.

□

2.4.3 Step 3

Theorem 2.4.3. $\mu^\star$ *is countably additive on* $\mathcal{L}$

Proof. First, observe that $\mu^\star(\varnothing) = \varnothing$ trivially, as $\varnothing \in \mathcal{J}$ and $\mu(\varnothing) = 0$. Now, assume $A, B \in \mathcal{L}$ with $A \subseteq B$. Note that for any $J \in \mathcal{J}$ with $B \subseteq J$, $A \subseteq J$. And thus,

$$\mu^\star(A) = \inf\{\mu(J) : A \subseteq J\} \leqslant \inf\{\mu(J) : B \subseteq J\} = \mu^\star(B)$$

It remains to verify that $\mu^\star$ is countably additive. Let us begin with finite additivity. It remains to show that for any $A, B \in \mathcal{L}$ that $\mu^\star(A \cup B) = \mu^\star(A) + \mu^\star(B)$. Fix $\epsilon > 0$. First, let $J_A, J_B \in \mathcal{J}$ be such that, $\mu(J_A) - \mu^\star(A) < \epsilon/2$ and likewise for J_B. It then follows that, $J_A \cup J_B$ is a valid cover of $A \cup B$, and so:

$$\mu^\star(A \cup B) \leqslant \mu(J_A \cup J_B) \leqslant \mu(J_A) + \mu(J_B) < \mu^\star(A) + \mu^\star(B) + \epsilon$$

Taking $\epsilon \to 0$, it's clear that $\mu^\star(A \cup B) \leqslant \mu^\star(A) + \mu^\star(B)$. Now, we would like the reverse inequality: $\mu^\star(A) + \mu^\star(B) \leqslant \mu^\star(A \cup B)$. To see this, suppose that J is such that $\mu(J) < \mu^\star(A \cup B) + \epsilon$. Consider any J_A, J_B s.t. $A \subseteq J_A$ and $B \subseteq J_B$. Now, let $J'_B = J_B \backslash J_A$. We still have $J'_B \supseteq B$. So then,

$$\mu^\star(A) + \mu^\star(B) < \mu(J_A \cap J) + \mu(J'_B \cap J)$$

By additivity on the ring and monotonicity,

$$= \mu((J_A \cap J) \cup (J'_B \cap J)) \leqslant \mu(J) \leqslant \mu^\star(A \cup B) + \epsilon$$

Taking $\epsilon \to 0$, we have proven finite additivity in the $n = 2$ case; the general finite case follows easily by induction.

Now, we proceed to countable additivity. Suppose that $A_1, A_2 ... \in \mathcal{L}$ are all disjoint. Let $A = \cup_i A_i$. First, observe by finite addivity and monotonicity that,

$$\mu^\star(A) \geqslant \mu^\star(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu^\star(A_i)$$

Taking $n \to \infty$, we have one inequality. It remains to show the opposite. Again, fix $\epsilon > 0$. We can this by showing that, for all $\epsilon > 0$,

$$\mu^\star(A) \leqslant \sum_{i=1}^{\infty} \mu^\star(A_i) + \epsilon$$

To see this, let J_i be such that $\mu(J_i) < \mu^\star(A_i) + \epsilon/2^i$. And let $A \subseteq J$. It follows that $A_i \subseteq J \cap J_i \subseteq J_i$, so $\mu(J \cap J_i) < \mu(A_i) + \epsilon/2^i$. And thus, by countable additivity on the ring $\mathcal{J}$, we obtain,

$$\mu^\star(A) \leqslant \mu(J) = \sum_{i=1}^{\infty} \mu(J \cap J_i) \leqslant \sum_{i=1}^{\infty} \mu^\star(A_i) + \epsilon/2^i = \sum_{i=1}^{\infty} \mu^\star(A_i) + \epsilon$$

Taking $\epsilon \to 0$, we conclude the desired result. $\square$

2.4.4 Step 4

Theorem 2.4.4. *$\mathcal{L}$ is a σ-algebra*

Proof. First, clearly $\varnothing \in \mathcal{L}$. Additionally, note that if $E \in \mathcal{L}$, then, for all $A \subseteq \mathbb{R}^d$, we have,

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

Which also implies that E^c is Lebesuge-measurable. Thus, $\mathcal{L}$ is closed under complement. Let us now assume that $E_1, E_2.. \in \mathcal{L}$. We then have that, letting $E = \cup_i E_i$,

$$\begin{aligned} \mu^\star(A) = \mu^\star(\cup_i A_i) = \sum_i \mu^\star(E_i) &= \sum_i \mu^\star(A \cap E_i) + \mu^\star(A \cap E_i^c) \\ &= \sum_i \mu^\star(A \cap E_i) + \sum_i \mu^\star(A \cap E_i^c) \\ = \mu^\star(\cup_i A \cap E_i) + \mu^\star(\cup_i A \cap E_i^c) &= \mu^\star(A \cap E) + \mu^\star(A \cap E^c) \end{aligned}$$

And thus E is Lebesgue measurable. This concludes the proof. $\square$

2.4.5 Step 5

Now, we have that $J \subseteq \mathcal{L}$ and $\sigma(\mathcal{J}) = \mathcal{B}(\mathbb{R}^d)$. And so, $\mathcal{B}(\mathbb{R}^d) \subseteq \sigma(\mathcal{J}) \subseteq \sigma(\mathcal{L}) = \mathcal{L}$. And thus, $\mathcal{B}(\mathbb{R}^d) \subseteq \mathcal{L}$. And since $\mu^\star$ is a countably additive measure on $\mathcal{L}$, we find that $\mu^\star$ is also a countably additive measure on $\mathcal{B}(\mathbb{R}^d)$. This is the desired result.

2.5 The π-λ Theorem and Uniqueness

We will define two families of sets, state & prove the π-λ theorem, and give an application w.r.t. the Lebesgue measure.

Definition 11 (π-System)**.** *Say P is a π system if it is closed under intersection*

Definition 12 (λ-System)**.** *Say L is a λ system if,*

- $\varnothing \in L$
- *L is closed under complement*
- *If $A_1..A_n \in L$ and the A_i's are pairwise disjoint, then $\cup_{i=1}^n A_i \in L$*

Theorem 2.5.1 (The π-λ Theorem)**.** *Say P is a π system contained in a λ system L. Then $\sigma(P) \subseteq L$.*

Proof. We show that $\lambda(P)$, the smallest λ system containing P, is a σ algebra. Thus, $\sigma(P) \subseteq \lambda(P) \subseteq L$, since L is already a λ system. Thus, it remains to prove that $\lambda(P)$ is a σ algebra.

Proposition 10. *A family of sets which is a π and λ system is also a σ algebra.*

Proof. The closure properties of a σ algebra can be easily checked □

Thus, it remains to show that $\lambda(P)$ is a π-system, i.e. it is closed under intersection.

Lemma 2.5.2. *Let L be a λ system. For $A \in L$, let,*

$$L_A = \{B \in L : A \cap B \in L\}$$

Then L_A is a λ system.

Proof. Check the properties of a λ system □

Lemma 2.5.3. *The intersections of a λ system is a λ system*

Proof. Check the properties of a λ system (not hard) □

Consider the following set G:

$$G = \{A \in \lambda(P) \; s.t. \; A \cap E \in \lambda(P) \, , \forall E \in P\}$$

Obviously,

$$G = \bigcap_{E \in P} (\lambda(P))_E$$

Combining the above two lemmas, it follows that G is a λ system. As P is a π system, $P \subseteq G$, so $\lambda(P) \subseteq \lambda(G) \subseteq \lambda(P)$, and thus $\lambda(P) = G$. Thus,

$G = \lambda(P)$.

Now, we work out a little more. Write,

$$H = \{A \in \lambda(P) : A \cap B \in \lambda(P), \forall B \in \lambda(P)\}$$

Now we find that,

$$H = \cap_{A \in \lambda(P)} (\lambda(P))_E$$

And thus again, H is a λ system. We find that $\lambda(P) = H$. But obviously, H is a π system, so we are done.

□

We will now show that the restriction of $\mu^\star$ to $\mathcal{B}(\mathbb{R}^d)$ is the only thing we can do. Suppose there is a second measure μ' which respects μ over $\mathcal{J}$; we can show that $\mu' = \mu^\star$ over $\mathcal{B}(\mathbb{R}^d)$. With one assumption: assume $\mu^\star$ and μ' are σ-finite. How will this work? We proceed in the following steps:

- Let $\mathcal{D} = \{A \in \mathcal{B}(\mathbb{R}^d) : \mu^\star(A) = \mu'(A)\}$
- Argue $\mathcal{D}$ is a σ algebra.
- Observe $\mathcal{J} \subseteq \mathcal{D}$
- Deduce $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{J}) \subseteq \sigma(\mathcal{D}) = \mathcal{D}$
- Conclude $\mu^\star = \mu'$ over all of $\mathcal{B}(\mathbb{R}^d)$.

The only real work here is to show that $\mathcal{D}$ is in fact a σ algebra, as theother steps are self explanatory. Because μ is σ-finite, let us write that $\Omega = \cup_j B_j$, where $B_j \in \mathcal{J}$ is countable and $\mu(B_j) < \infty$ for all j.

We first do the proof for finite measures. First, note that we can consider the λ system $\mathcal{I}$. Clearly, if $\mu^\star = \mu'$ over $\mathcal{J}$, the same holds true over J. Now, we show $\mathcal{D}$ is a λ system. Clearly, $\varnothing \in \mathcal{D}$. Furthermore, $\mathcal{D}$ is closed under complement, since,

$$\mu^\star(A^c) = \mu^\star(\Omega) - \mu^\star(A)$$
$$= \mu'(\Omega) - \mu'(A) = \mu'(A^c)$$

As $\mu', \mu^\star$ are countably (and thus finitely) additive, L is obviously closed under disjoint unions as well, as,

$$\mu^\star(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mu^\star(A_i) = \sum_{i=1}^n \mu'(A_i) = \mu'(\cup_{i=1}^n A_i)$$

Thus, $\mathcal{D}$ is a λ system. We conclude from the π lambda theorem that $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{I}) \subseteq \mathcal{D}$, so $\mu^\star = \mu'$ over $\mathcal{B}(\mathbb{R}^d)$.

The general σ-finite is simple. Let $B_1, B_2..$ be disjoint in $\mathcal{J}$ with $\cup_i B_i = \Omega$ and $\mu'_i(B_i) = \mu^\star(B) < \infty$. Then define $\mu^\star_i(A) = \mu^\star(A \cap B_i), \mu'_i(A) = \mu'(A \cap B_i)$ for al $A \in \mathcal{B}(\mathbb{R}^d)$. It follows that,

$$\mu' = \sum_i \mu'_i \qquad \mu^\star = \sum_i \mu^\star_i$$

And since each $\mu'_i, \mu^\star_i$ is a finite measure, by our prior work, they must agree. And so, all of $\mu', \mu^\star$ agree. This proves the unqiqueness, as desired!

2.6 Consequences

We should pat ourselves on the back and say.... whew. We are basically done with the hard work. For example, we can now define probability measures to our heart's content! For example, if we say,

$$\mu((a_1, b_1]... \times (a_d, b_d]) = \int_{a_1}^{b_1} .. \int_{a_d}^{b_d} \left(\frac{1}{2\pi}\right)^{-d/2} \prod_{i=1}^{d} e^{-\frac{1}{2}x_i^2} dx_i$$

Then we know μ is the unique measure on $\mathcal{B}(\mathbb{R}^d)$ corresponding to the normal distribution!

Chapter 3

Measurable Functions and Integrating Functions

We will now devote ourselves to the study of functions acting on measure spaces. First, we define what it means for a function to be measurable. Then, we build up our study of how to integrate functions. Throughout the chapter, we let $(\Omega, \mathcal{F}, \mu)$ be a fixed measure space.

3.1 Measurable Functions

3.1.1 Intuition

Let us return to chapter 1. There, we said that Ω could be an incredibly rich universe of possible events, but $\mathcal{F}$ is a subset of interest. For example, in a dice roll, Ω could differentiate the outcomes of the dice roll, the weather tomorrow, and what you eat for lunch tomorrow. But we may let $\mathcal{F} = \sigma(\cup_{i=1}^{6}\{\omega : \text{dice roll at } \omega \text{ is } i\})$ be the σ-algebra which captures enough richness for our purposes. We think of measurable functions as those which compose well with the set of events we care about. This is a sort of necessary but sufficient condition for our study of probability.

Example 4. *Let* $\Omega = \{1, 2, 3, 4, 5, 6\}, \mathcal{F} = \sigma(\{\{1, 2\}, \{3, 4\}, \{5, 6\}\})$ *and* $\mu(\{1, 2\}) = \mu(\{3, 4\}) = \mu(\{5, 6\}) = 1/3$*. Finally, let* $X(\omega) = 2\omega$*.*

In this example, X is not measurable in some sense. Think about it. What is the probability that $X = 2$? Our measure is underspecified! Just because we know $\mu(X \in \{2, 4\}) = 1/3$ doesn't mean we can determine $\mu(\{X \in 2\}) = \mu(\{1\})$. It could be that we have a biased die in which $\mu(\{1\}) = 3/24, \mu(\{2\}) = 1/24$. So what is the expected value of X? There really isn't enough information! On the other hand, if $X = 1$ if $\omega \in \{1, 2, 3, 4\}$ and is 0 otherwise, we can determine the expected value of X, because $X^{-1}(\{1\}) = \{1, 2, 3, 4\} \in \mathcal{F}$; likewise, $X^{-1}(\{0\}) = \{5, 6\} \in \mathcal{F}$.

3.1.2 Definition of Measurability

Suppose $(\Omega, \mathcal{F}, \mu)$ is a measure space and $(\mathcal{X}, \mathcal{A})$ is a family of sets equipped with a σ-algebra.

Definition 13. *A function $X : \Omega \to \mathcal{X}$ is said to be $\mathcal{F}/\mathcal{A}$ measurable, or simply measurable, if for all $A \in \mathcal{A}$, $X^{-1}(A) \in \mathcal{F}$. One may write $X : (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A})$.*

Definition 14. *If $\mathcal{X} = \mathbb{R}, \mathcal{A} = \mathcal{B}(\mathbb{R})$, the set of all $\mathcal{F}/\mathcal{B}(\mathbb{R})$ measurable functions is denoted $\mathcal{M}(\Omega, \mathcal{B}(\mathbb{R}))$. The set of all such nonnegative functions is $\mathcal{M}^+(\Omega, \mathcal{B}(\mathbb{R}))$.*

3.1.3 Determining Measurability

Certainly, we could check 13 by simply taking arbitrary elements of $\mathcal{A}$ and checking $X^{-1}(\mathcal{A}) \in \mathcal{F}$. This may prove to be a difficult task, however. For example, recall from our previous study of the Lebesgue measure that $\mathcal{B}(\mathbb{R})$ is complicated! We will show that it suffices to check a generating class.

Theorem 3.1.1. *If $\mathcal{A} = \sigma(\mathcal{E})$ and for all $E \in \mathcal{E}$, $X^{-1}(E) \in \mathcal{F}$, then X is $\mathcal{F}/\mathcal{A}$ measurable.*

Proof. We proceed via a generating class argument. Let,

$$\mathcal{D} = \{A \in \mathcal{A} : X^{-1}(A) \in \mathcal{F}\}$$

We shall show that $\mathcal{D}$ is a σ algebra. First, observe that $X^{-1}(\emptyset) = \emptyset \in \mathcal{D}$. Furthermore, if $A \in \mathcal{D}$, as $X^{-1}(A^c) = X^{-1}(A)^c$, $X^{-1}(A) \in \mathcal{F}$, and $\mathcal{F}$ is closed under complement, $A^c \in \mathcal{D}$. Thus $\mathcal{D}$ is closed under complement. Finally, suppose $A_1, A_2... \in \mathcal{D}$. One can check that $X^{-1}\big(\cup_{i=1}^{\infty} A_i\big) = \cup_{i=1}^{\infty} X^{-1}(A_i)$. And as each $X^{-1}(A_i) \in \mathcal{F}$, the whole countable union is as well. Thus, $\cup_{i=1}^{\infty} A_i \in \mathcal{D}$. We conclude that $\mathcal{D}$ has the desirable closure properties of a σ algebra.

By assumption, $\mathcal{E} \subseteq \mathcal{D}$. Therefore, $\mathcal{A} = \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{D}) = \mathcal{D}$. Also, $\mathcal{D} \subseteq \mathcal{A}$ by definition, so $\mathcal{A} = \mathcal{D}$. We conclude that X is measurable. $\square$

3.1.4 Examples

If $\mathcal{F} = \mathcal{B}(\mathbb{R}^n), \mathcal{A} = \mathcal{B}(\mathbb{R}^d)$ and X is continuous, then X is measurable. It is a standard fact that continuous functions map open sets to open sets, and the preimage of an open set is open. Thus, let $\mathcal{G}_n$ be the open sets of $\mathbb{R}^n$ and $\mathcal{G}_d$ be the open sets of $\mathbb{R}^d$. Clearly, by this fact, for each $E \in \mathcal{G}_d, X^{-1}(E) \in \mathcal{G}_n$. And since $\mathcal{G}_d$ generates $\mathcal{B}(\mathbb{R}^d)$, by theorem 3.1.1, X is measurable.

3.1.5 Properties

Theorem 3.1.2. *If $X : (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{A})$ and $Y : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$, then $Y \circ X : \Omega \to \mathcal{Y}$ is $\mathcal{F}/\mathcal{B}$ measurable.*

Proof. This is perfectly straightforward to check using the definitions. ◻

Theorem 3.1.3. *If $X, Y \in \mathcal{M}(\Omega, \mathcal{F})$ are both bounded and measurable, then so is $X + Y$ and XY*

Proof. First, define $T = X + Y$. First, observe for any interval (a, b),

$$T^{-1}(a,b) = \{\omega : X(\omega) + Y(\omega) > a\} \cap \{\omega : X(\omega) + Y(\omega) < b\}$$

Note that,

$$\begin{aligned}\{\omega : X(\omega) + Y(\omega) > a\} &= \bigcup_{q \in \mathbb{Q}} \{\omega : X(\omega) > q\} \cap \{\omega : Y(\omega) > a - q\} \\ &= \bigcup_{q \in \mathbb{Q}} X^{-1}(q, \infty) \cap Y^{-1}(a - q, \infty)\end{aligned}$$

Note that $X^{-1}(q, \infty) \in \mathcal{F}, Y^{-1}(a - q, \infty) \in F$ by measurability. And by closure properties of σ algebras, the above is in $\mathcal{F}$. Likewise, $\{\omega : X(\omega) + Y(\omega) < b\} \in \mathcal{F}$. Again, by closure under intersection, $T^{-1}(a, b) \in \mathcal{F}$. Since the open intervals generate $\mathcal{B}(\mathbb{R})$, this is sufficient for measurability by theorem 3.1.1.

Now, we show XY is measurable in a somewhat similar fashion. We proceed like so:

- Consider the map $T(\omega) = (X(\omega), Y(\omega))$ and $\psi(u, v) = uv$. Note $XY = \psi \circ T$.
- As $\{A \times B : A \in \mathcal{B}(\mathbb{R}), B \in \mathcal{B}(\mathbb{R})\}$ generates $\mathcal{B}(\mathbb{R}^2)$, and X and Y are measurable, T is measurable. Why? Because, for $A, B \in \mathcal{B}(\mathbb{R})$,

  $$T^{-1}(A \times B) = \underbrace{X^{-1}(A)}_{\in \mathcal{F}} \cap \underbrace{Y^{-1}(B)}_{\in \mathcal{F}} \in \mathcal{F}$$

  So by theorem 3.1.1, this is sufficient to say that T is measurable.
- As $\psi : \mathbb{R}^2 \to \mathbb{R}$ is continuous, it is measurable
- Thus, XY can be regarded as the composition of measurable functions, so it is measurable.

◻

Theorem 3.1.4. *If $X_1, X_2..$ is a sequence of measurable functions, then $X = \sup_i X_i$ is measurable.*

Proof. Note intervals of the form (a, ∞) generate $\mathcal{B}(\mathbb{R})$. Furthermore,

$$X^{-1}(a, \infty) = \{\omega : X(\omega) > a\} = \bigcup_{i=1}^{n} X_i^{-1}(a, \infty)$$

Thus, $X^{-1}(a, \infty) \in \mathcal{F}$. Since these generate $\mathcal{B}(\mathbb{R})$, we are done. ◻

Theorem 3.1.5. *If $X_1, X_2..$ is a sequence of measurable functions, then $X = \inf_i X_i$ is measurable.*

Proof. The proof is analogous. □

Proposition 11. *The* lim sup *and* lim inf *of measurable functions is measurable.*

Proof. We will show the proof for the lim sup case as the lim inf case is analogous. Recall if $X_1, X_2...$ are measurable functions, then if $X = \limsup_i X_i$,

$$X(\omega) = \inf_n \sup_{m \geqslant n} X_i(\omega)$$

As $\sup_{m \geqslant n} X_i(\omega)$ is measurable for each n, and the infimum of measurable functions is measurable, X is measurable. □

3.2 The Integral

We will now develop the notion of integrals of functions, from the ground up. First, we begin with a measure space $(\Omega, \mathcal{F}, \mu)$. Think of integrals as functionals: maps from the space of measurable functions to $\mathbb{R}$. While notation varies, we will adopt two ways of denoting the integral of a function X with respect to a measure μ:

$$\int X d\mu \qquad \text{and} \qquad \mu(X)$$

While the $d\mu$ does relate to the dx from Riemannian integration, ignore this for now. Think of the $d\mu$ merely as a symbol which says we are integrating with respect to μ, rather than some other measure.

3.2.1 Simple Functions

A simple function will be of the form,

$$X(\omega) = \sum_{i=1}^{n} \alpha_i \mathbb{I}\{A_i\}$$

Where each $\alpha_i \geqslant 0$ and $A_i \in \mathcal{F}$. For such an X, we will define its integral like so:

$$\int X d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i)$$

If $\mu(A_i) = \infty, \alpha_i = 0$, adopt the convention that $\alpha_i \mu(A_i) = 0$ — this is the natural thing to do, as we don't want our integral to depend on sets which don't contribute to our function. It remains to verify consistency.

Proposition 12. *Suppose that X can be written as $X = \sum_{j=1}^{m} \beta_j \mathbb{I}\{B_j\} = \sum_{i=1}^{n} \alpha_i \mathbb{I}\{A_i\}$. Then, $\sum_{i=1}^{n} \alpha_i \mu(A_i) = \sum_{j=1}^{m} \beta_j \mu(B_j)$. And so, the integral of a simple function is a well-defined object.*

Proof. Assume without loss of generality that $\cup_{j=1}^{m} B_j = \cup_{i=1}^{n} = A_i$. Otherwise, we could simply consider $\sum_{i=1}^{n} \alpha_i \mathbb{I}\{A_i\} + 0 \cdot \mathbb{I}\{\Omega - \cup_{i=1}^{n} A_i\}$ without changing X or its integral. Furthermore, assume without loss of generality that the A_i's are disjoint, as are the B_j's. Let $\gamma_{i,j} = X(\omega)$ for $\omega \in A_i \cap B_j$.First, note that X can be written as:

$$X = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \mathbb{I}\{A_i \cap B_j\} = \sum_{j=1}^{m} \sum_{i=1}^{n} \beta_j \mathbb{I}\{B_j \cap A_i\} = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{i,j} \mathbb{I}\{A_i \cap B_j\}$$

Note for fixed $i, \gamma_{i,j} = \alpha_i$. For fixed j, $\gamma_{i,j} = \beta_j$. And so,

$$\sum_{i=1}^{n} \alpha_i \mu(A_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \mu(A_i \cap B_j)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{i,j} \mu(A_i \cap B_j) = \sum_{j=1}^{m} \sum_{i=1}^{n} \gamma_{i,j} \mu(A_i \cap B_j)$$

$$= \sum_{j=1}^{m} \sum_{i=1}^{n} \beta_j \mu(A_i \cap B_j) = \sum_{j=1}^{m} \beta_j \mu(B_j)$$

□

3.2.2 Properties of Integrals of Simple Functions

Proposition 13. *$\int \alpha X + \beta Y d\mu = \alpha \int X d\mu + \beta \int Y d\mu$. Thus, the integral is a linear functional.*

Proof. Let,

$$X = \sum_{i=1}^{n} \alpha_i \mathbb{I}\{A_i\}, Y = \sum_{j=1}^{m} \beta_j \mathbb{I}\{B_j\}$$

Again, assume without loss of generality that the A_i's are pairwise disjoint and span Ω. Then,

$$\alpha X + \beta Y = \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha \alpha_i + \beta \beta_j) \mathbb{I}\{A_i \cap B_j\}$$

Thus,

$$\int \alpha X + \beta Y d\mu = \sum_{i=1}^{n} \sum_{j=1}^{m} (\alpha \alpha_i + \beta \beta_j) \mu(A_i \cap B_j)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha\alpha_i \mu(A_i \cap B_j) + \sum_{j=1}^{m} \sum_{i=1}^{n} \beta\beta_j \mu(A_i \cap B_j)$$

$$= \alpha \sum_{i=1}^{n} \alpha_i \mu(A_i) + \beta \sum_{j=1}^{m} \beta_j \mu(B_j) = \alpha \int X d\mu + \beta \int Y d\mu$$

□

3.2.3 Extension to All Measurable Functions

For a general measurable function $X \in \mathcal{M}^+(\Omega, \mathcal{F})$, we define its integral to be the highest among those simple functions which are less than X:

$$\int X d\mu = \sup_{s_n \leqslant X \text{ simple}} \int s_n d\mu$$

Theorem 3.2.1 (This Integral is Nice)**.** *Suppose X and Y are measurable functions. Then,*

1. $\int \alpha X + \beta Y d\mu = \alpha \int X d\mu + \beta \int Y d\mu$
2. *If $X = Y$ almost everywhere, then $\int X d\mu = \int Y d\mu$.*
3. *If $X \leqslant Y$ almost everywhere, then $\int X d\mu \leqslant \int Y d\mu$*

Proof. **Proof of 1**: Let x_n, y_n be simple functions such that $\int X d\mu < \int x_n d\mu + 1/(2\alpha n)$, $\int Y d\mu < \int y_n d\mu + 1/(2\beta n)$. Then, note $\alpha X + \beta Y \geqslant \alpha x_n + \beta y_n$. So,

$$\int \alpha X + \beta Y d\mu \leqslant \int \alpha x_n + \beta y_n d\mu = \alpha \int x_n d\mu + \beta \int y_n d\mu$$

$$\leqslant \alpha\Big(\int X d\mu + 1/(2\alpha n)\Big) + \beta\Big(\int Y d\mu + 1/(2\beta n)\Big)$$

$$= \alpha \int X d\mu + \beta \int Y d\mu + 1/n$$

Taking $n \to \infty$, we have $\int \alpha X + \beta Y d\mu \leqslant \alpha \int X d\mu + \beta \int Y d\mu$. For the reverse inequality,

$$\int \alpha X + \beta Y d\mu = \sup_n \Big\{ \int z_n d\mu : z_n \leqslant \alpha X + \beta Y \Big\}$$

$$\geqslant \sup_n \Big\{ \int \alpha x_n + \beta y_n d\mu : x_n \leqslant X, y_n \leqslant Y \Big\}$$

Since x_n, y_n can vary freely,

$$= \alpha \sup_n \Big\{ \int x_n d\mu : x_n \leqslant X \Big\} + \beta \sup_n \Big\{ \int y_n d\mu : y_n \leqslant Y \Big\}$$

$$= \alpha \int X d\mu + \beta \int Y d\mu$$

Which gives the other side of the inequality. So we are done.

Proof of 2: Let $N = \{\omega : X(\omega) \neq Y(\omega)\}$. By assumption, $\mu(N) = 0$, so N is negligible. Let $\tilde{X} = X\mathbb{I}\{N^c\}$. We show that $\int X d\mu = \int \tilde{X} d\mu$. It then follows from symmetry and the fact that $\tilde{Y} = \tilde{X}$ that the desired result is true. Note for any simple function $x = \sum_{i=1}^{n} \alpha_i \mathbb{I}\{A_i\}$,

$$\int x d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i) = \sum_{i=1}^{n} \alpha_i \mu(A_i \cap N^c) = \int x\mathbb{I}\{N^c\} d\mu$$

Thus, taking supremums,

$$\int X d\mu = \sup \left\{ \int x d\mu : x \leqslant X \right\} = \sup \left\{ \int x\mathbb{I}\{N^c\} d\mu : x \leqslant X \right\}$$

$$= \sup \left\{ \int x d\mu : x \leqslant X\mathbb{I}\{N^c\} \right\} = \int \tilde{Xd}\mu$$

And thus,

$$\int X d\mu = \int \tilde{X} d\mu = \int \tilde{Y} d\mu = \int Y d\mu$$

Proof of 3: First, suppose $X \leqslant Y$ everywhere. The property will hold by definition of supremum. Note since $X \leqslant Y$,

$$\int X d\mu = \sup \left\{ \int x : x \leqslant X \right\} \leqslant \sup \left\{ \int x : x \leqslant Y \right\} = \int Y d\mu$$

Now, if $X > Y$ on a negligible set N. Consider, $\tilde{X} = X\mathbb{I}\{N^c\}$. Then $\int \tilde{X} d\mu = \int X d\mu$ by property 2. And $\int \tilde{X} d\mu \leqslant \int Y d\mu$ by our earlier work, so we are done! ☐

Definition 15 (Convergence Definitions)**.** *We use the following conventions from here on:*

- *Numbers: Say $a_1, a_2 ... \uparrow a$ if $a_1 \leqslant a_2 ...$ and $\lim_{n\to\infty} a_n = a$*
- *Sets: Say $A_1, A_2 \uparrow A$ if $A_1 \subseteq A_2 \subseteq A_3 ...$ and $\cup_{i=1}^{\infty} A_i = A$*
- *Sets: Say $A_1, A_2 \downarrow A$ if $A_1 \supseteq A_2 \supseteq A_3 ...$ and $\cap_{i=1}^{\infty} A_i = A$*
- *Functions: Say $X_n \uparrow X$ if $X_n \leqslant X_{n+1}$ for all n and $X_n \to X$ pointwise*

The monotone convergence theorem is a fundamental result in the theory of integration. To prove it, we will use the so called continuity of a measure:

Theorem 3.2.2. *If $A_n \downarrow \varnothing$ with at least one $\mu(A_i) < \infty$, then $\mu(A_n) \downarrow 0$. Also, if $A_n \uparrow A$, then $\mu(A_n) \uparrow \mu(A)$.*

Proof. Assume WLOG that $\mu(A_1) < \infty$. Let $B_1 = A_1 \backslash A_2, B_2 = A_2 \backslash A_3$, and so on: $B_i = A_{i-1} \backslash A_i$. It's clear that $\cup_i B_i = A_1$. Thus, by countable additivity,

$$\mu(A_1) = \sum_{i=1}^{\infty} \mu(B_i)$$

On the other hand,

$$\mu(A_1) = \sum_{i=1}^{n} \mu(B_i) + \sum_{i=n+1}^{\infty} \mu(B_i) = \sum_{i=1}^{n} \mu(B_i) + \mu(A_n)$$

Taking $n \to \infty$, it must be that $\sum_{i=1}^{n} \mu(B_i) \to \mu(A_1)$. And thus, for equality to hold, $\mu(A_n) \downarrow 0$. The proof of the second statement works in much the same way. $\square$

Theorem 3.2.3 (The Monotone Convergence Theorem)**.** *If $X_n \uparrow X$ pointwise almost everywhere, then X is measurable and $\int X_n d\mu \uparrow \int X d\mu$*

Proof. First, assume without loss of generality that $X_n \uparrow X$ everywhere, as theorem 3.2.1 would readily imply the desired result. Also, the measurability of X has already been proven, as it is the sup of measurable functions. Furthemore, for any n, $\int X_n d\mu \leqslant \int X d\mu$ as $X_n \leqslant X$. So taking $n \to \infty$, we find $\lim_n \int X_n d\mu \leqslant \int X d\mu$. It remains to show that $\lim_n \int X_n d\mu \geqslant \int X d\mu$. Indeed, let X_m be a sequence of simple functions such that $\int X_m d\mu > \int X - 1/m$. Now define $X_{n,m} = X_m(1 - \frac{1}{m})\mathbb{I}\{X_n \geqslant (1 - \frac{1}{m}X_m)\}$. Note that $X_{n,m}$ is simple, and $X_{n,m} \leqslant X_n$. And thus,

$$\int X_{n,m} d\mu \leqslant \int X_n d\mu$$

Now assume that $X_m = \sum_{i=1}^{n_m} \alpha_{i,m} \mathbb{I}\{A_{i,m}\}$ so that $X_{n,m} = (1-\frac{1}{m}) \sum_{i=1}^{n_m} \alpha_{i,m} \mathbb{I}\{A_{i,m} \cap \{X_n \geqslant (1 - \frac{1}{m}X_m)\}\}$. Thus,

$$\int X_n d\mu \geqslant (1 - \frac{1}{m}) \sum_{i=1}^{n_m} \alpha_{i,m} \mu(\{A_{i,m} \cap \{X_n \geqslant (1 - \frac{1}{m}X_m)\}\})$$

Taking $n \to \infty$, the continuity of measures implies that $A_{i,m} \cap \{X_n \geqslant (1 - \frac{1}{m}X_m)\} \uparrow A_{i,m}$. Thus,

$$\lim_{n \to \infty} \int X_n d\mu \geqslant (1 - \frac{1}{m}) \int X_m d\mu \geqslant (1 - \frac{1}{m}) \left(\int X d\mu - \frac{1}{m} \right)$$

Taking $m \to \infty$, we are done!

$\square$

3.2.4 Integrals of More General Functions

Our last iteration of building the integral is that for possibly negative functions. For $X \in \mathcal{M}(\Omega, \mathcal{F})$, let $X^+ = X\mathbb{I}\{X \geqslant 0\}$ and $X^- = |X|\mathbb{I}\{X \leqslant 0\}$. It follows that $X = X^+ - X^-$. If at least one of $\int X^+ d\mu, \int X^- d\mu$ is finite, we define $\int X d\mu = \int X^+ d\mu - \int X^- d\mu$. Otherwise, the integral is not defined. If the quantity is finite, the function is said to be integrable. One can easily verify the same niceness properties as before by splitting arbitrary functions into their positive and negative parts.

3.3 Limit Theorems

3.3.1 Fatou's Lemma

Fatou's Lemma is a relatively simple theorem to prove which will allow us to prove many limit results down the road.

Lemma 3.3.1 (Fatou's Lemma). *For* $\{X_n\}_n \in \mathcal{M}^+(\Omega, \mathcal{F})$, $\int \liminf_n X_n d\mu \leqslant \liminf \int X_n d\mu$.

Proof. First, to remember this inequality, think of the lim inf of a function as containing many more degrees of freedom than the lim inf of the integral, and thus we are lower. Note first that, for all n, $\inf_n X_n \leqslant X_n$. Define $Y_m = \inf_{n \geqslant m} X_n$. The monotone convergence theorem implies,

$$\lim_{m \to \infty} \int Y_m = \int \sup Y_m$$

And thus, substituting our definitions,

$$\lim_{m \to \infty} \int \inf_{n \geqslant m} X_n d\mu = \int \sup_m \inf_{n \geqslant m} X_n = \int \liminf_n X_n d\mu$$

Note also that for all m, $Y_m \leqslant X_m$. Thus,

$$\liminf_m \int Y_m d\mu \leqslant \liminf_m \int X_m d\mu$$

But when a limit exists, it is equal to the lim inf, so,

$$\int \liminf_n X_n d\mu = \liminf_m \int Y_m d\mu \leqslant \liminf_m \int X_m d\mu$$

And we're done!

□

3.3.2 The Dominated Convergence Theorem

Next comes one of the most important results yet! It comes up time and time again and is arguably the most general tool or interchanging limits.

Theorem 3.3.2 (The Dominated Convergence Theorem)**.** *Let X_n converge pointwise to a function X almost everywhere. Assume $|X_n| \leqslant Y$ for some integrable function Y. Then $\lim_n \int X_n d\mu = \int X d\mu$.*

Proof. As Y is integrable, each X_n is integrable. Furthermore, $Y - X_n, Y + X_n$ are integrable and nonnegative, by the tringle inequality. Therefore,

$$\int \liminf_n (Y - X_n) d\mu \leqslant \liminf_n \int (Y - X_n) d\mu$$

$$\int \liminf_n (Y + X_n) d\mu \leqslant \liminf_n \int (Y + X_n) d\mu$$

By linearity of the integral and the lim inf, this implies,

$$\int \liminf_n (-X_n) d\mu \leqslant \liminf_n \left(- \int X_n d\mu \right)$$

$$\int \liminf_n X_n d\mu \leqslant \liminf_n \int X_n d\mu$$

Note the first line is equivalent to,

$$\int \limsup_n X_n d\mu \geqslant \limsup_n \int X_n d\mu$$

But we assumed that X_n converges, so $\liminf_n X_n = \limsup_n X_n = X$. Combining, we obtain,

$$\limsup_n \int_n X_n d\mu \leqslant \int X d\mu \leqslant \liminf_n \int X_n d\mu$$

But obviously, $\limsup_n \int_n X_n \geqslant \liminf_n \int_n X_n$. This forces their equality, and thus the assertion follows as claimed.

3.3.3 Application: Differentiation Under the Integral

Oftentimes, you will see something of the essence of:

$$\frac{\partial}{\partial t} \int f(x,t) dx = \int \frac{\partial}{\partial t} f(x,t) dx$$

How could this be true? It could be easily understood as an application of the dominated convergence theorem. Indeed, consider for some n,

$$X_{n,t}(x) = n(f(x, t + 1/n) - f(x,t))$$

It is clear to see that, for fixed x, $\lim_{n\to\infty} X_n(x,t) = \frac{\partial}{\partial t} f(x,t)$. Thus, the tempting interchange of limits is:

$$\int \frac{\partial}{\partial t} f(x,t)dx = \int \lim_n X_{n,t} dx = \lim_n \int X_{n,t} dx = \lim_n \int X_{n,t} dx$$
$$= \lim_n n\left(\int f(x,t+1/n)dx - \int f(x,t)dx \right) = \frac{\partial}{\partial t} \int f(x,t)dx$$

How do we make this rigorous? Well first obesrve that, in a way, this is a claim about differentiation at each t. So fix t. For dominated convergence, we require that $X_{n,t}$ be dominated by an integrable function $Y_t(x)$, where this function is allowed to depend on t, (but not on n — so there is some dependence on t). If f is continuously differentiable with respect to t, it suffices to bound $\frac{\partial}{\partial t} f(x,t)$ at t, since then this can be extended to a more local bound. Of course, nothing is stopping us from collecting our bounding functions into a function of two variables. Summarizing our analysis into a theorem, we have:

Theorem 3.3.3. *Let $f(x,t) : \mathbb{R}^2 \to \mathbb{R}$ be a function which is continuous differentiable in an open set with respect to t. Suppose there exists a function $F(x,t)$ which is integrable for fixed x and $|\frac{\partial}{\partial t} f(x,t)| \leqslant F(x,t)$. Then, $\frac{\partial}{\partial t} \int f(x,t)dx = \int \frac{\partial}{\partial t} f(x,t)dx$ for all t.*

□

3.4 The Borel Cantelli Lemma

This is a bit of an aside about a probabilistic technique that will come up often. It works very simply.

Definition 16. *As a matter of notation, for a countable family of sets $A_1, A_2 ...$, define,*

$$\{A_n \ i.o.\} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega : \omega \text{ in infinitely many } A_n\}$$

This set can be understood as those events which occur "infinitely" often in the sequence A_1, A_2..

Lemma 3.4.1 (The Borel Cantelli Lemma (Part 1))**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, $\mathbb{P}(\{A_n \ i.o.\}) = 0$.*

Proof. Note if $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A_n) \to 0$. Additionally, $\lim_n \sum_{m=n}^{\infty} \mathbb{P}(A_m) \to 0$. And thus, for any n,

$$\mathbb{P}\{A_n \ i.o.\} \leqslant \mathbb{P}\{\bigcup_{m=n}^{\infty} A_m\} \leqslant \sum_{m=n}^{\infty} \mathbb{P}(A_m)$$

Taking $n \to \infty$, the right hand side becomes 0. This proves part 1 of the lemma. We will be able to prove a converse in a Part 2 once we introduce notions of independence. □

3.5 Special Case: Expectations

When $\mu = \mathbb{P}$ is a probability measure, the corresponding integral corresponds to an expectation. In this, we write,

$$\int X d\mathbb{P} = \mathbb{E}[X]$$

We will typically adopt the latter when the measure is unambiguous.

3.6 Convexity

3.6.1 Convex Combinations

Recall that a convex function $f : U \to V$, where $U, V \subseteq \mathbb{R}$, has the property that for all $x, y \in U, \lambda \in [0,1]$,

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y)$$

Note that λ and $1-\lambda$ provide a so called *convex combination* of numbers. In general, $a_1, a_2 ... a_n$ is called a convex combination if $a_i > 0$ for all i and $\sum_{i=1}^n a_i = 1$. One can easily verify that the above relation holds for more general convex combinations:

Lemma 3.6.1. *Let $a_1..a_n$ be a convex combination with $x_1, x_2...x_n \in U$. Then, $f(\sum_{i=1}^n a_i x_i) \leqslant \sum_{i=1}^n a_i f(x_i)$.*

Proof. We proceed by induction. The $n = 2$ case is the direct definition of convexity. Now suppose the $n-1$ case is true. First, suppose $0 < a_n < 1$; otherwise, the inductive step is trivial. We have,

$$f(\sum_{i=1}^{n} a_i x_i) = f(\sum_{i=1}^{n-1} a_i x_i + a_n x_n)$$

$$= f((1-a_n) \sum_{i=1}^{n-1} \frac{a_i}{1-a_n} x_i + a_n x_n)$$

We recognize a_n and $1-a_n$ as a convex combination, and so,

$$\leqslant a_n f(x_n) + (1-a_n) f(\sum_{i=1}^{n-1} \frac{a_i}{1-a_n} x_i)$$

We now recognize $\frac{a_1}{1-a_n} ... \frac{a_{n-1}}{1-a_n}$ as a convex combination. And so,

$$\leqslant a_n f(x_n) + (1-a_n) \sum_{i=1}^{n-1} \frac{a_i}{1-a_n} f(x_i) = \sum_{i=1}^{n} a_i f(x_i)$$

This proves the lemma.

$\square$

Corollary 3.6.1.1. *If $a_1..a_n$ is a convex combination and $x_1...x_n > 0$, then,*

$$\log(a_1x_1 + ..a_nx_n) \geqslant a_1 \log(x_1) + ... + a_n \log(x_n)$$

Proof. This follows from the convexity of $-\log$. $\square$

Another fact about convex functions, which we will not prove, is that they are continuous.

3.6.2 Jensen's Inequality

Note that there is some sense in which a probability distribution provides a sort of generalized convex combination. From the above, it follows that if X is a discrete random variable with state space $\Omega = \{x_1...x_n\}$, then:

$$\mathbb{E}[f(X)] = \sum_{i=1}^{n} f(x_i)\mathbb{P}\{X = x_i\}$$

$$f(\mathbb{E}(X)) = f\left(\sum_{i=1}^{n} x_i\mathbb{P}\{X = x_i\}\right)$$

Note that we recognize $\mathbb{P}\{X = x_1\}...\mathbb{P}\{X = x_n\}$ as a convex combination. And thus,

$$f(\mathbb{E}[X]) \leqslant \mathbb{E}[f(X)]$$

The same is true more generally.

Theorem 3.6.2 (Jensen's Inequality)**.** *Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ be a probability space and f be a convex function. Then, $f(\mathbb{E}[X]) \leqslant \mathbb{E}[f(X)]$.*

Proof. Bear in mind the finite case provided earlier. Intuiviely, you might imagine that the more general case is true by continuity. But to offer a reasonably concise proof, we will use another property of convex functions. A convex function f has, at each x, a support line ℓ_x such that $\ell_x(x) = f(x)$, and $\ell_x \leqslant f$ everywhere. So let $x_0 = \mathbb{E}[X]$ and let ℓ be the support line at x_0. If we let $\ell(x) = ax+b$, it's clear that $\mathbb{E}[\ell(X)] = a\mathbb{E}[X]+b = \ell(x_0)$. But by the decreasing property, $\mathbb{E}[\ell(X)] \leqslant \mathbb{E}[f(X)]$. Putting this all together, we find,

$$f(\mathbb{E}[X]) = f(x_0) = \ell(x_0) = \mathbb{E}[\ell(X)] \leqslant \mathbb{E}[f(X)]$$

$\square$

3.6.3 Hölder's Inequality

Conventional proofs of Hölder's Inequality typically rely on Young's Inequality. While these are generally faster, I think they are slightly less intuitive. Let us actually prove a general statement, from which Hölder's Inequality is a corollary.

Theorem 3.6.3. *Let $X_1 ... X_n$ be nonnegative measurable functions and $a_1 ... a_n$ be a convex combination. Then,*

$$\int \left(\prod_{i=1}^{n} X_i^{a_i} \right) d\mu \leqslant \prod_{i=1}^{n} \left(\int X_i d\mu \right)^{a_i}$$

Proof. It is simple to check that if $\left(\int X_i d\mu \right) = 0$ or ∞ for any i, then the corresponding inequality is trivial. Otherwise, assume all relevant integrals are finite and nonzero. We will first reduce the inequality to something simpler

$$\int \prod_{i=1}^{n} \left(\frac{X_i}{\int X_i d\mu} \right)^{a_i} d\mu \leqslant 1$$

So define $Y_i = \frac{X_i}{\int X_i d\mu}$. Clearly, each Y_i integrates to 1. It remains to show that $\int \prod_{i=1}^{n} Y_i^{a_i} d\mu \leqslant 1$. Indeed, we can check that, by the Corollary 3.6.1.1,

$$\prod_{i=1}^{n} Y_i^{a_i} = \exp\left(\log\left(\prod_{i=1}^{n} Y_i^{a_i} \right) \right) = \exp\left(\sum_{i=1}^{n} a_i \log(Y_i) \right)$$
$$\leqslant \exp\left(\log\left(\sum_{i=1}^{n} a_i Y_i \right) \right) = \sum_{i=1}^{n} a_i Y_i$$

The above is only rigorous when $\prod_{i=1}^{n} Y_i^{a_i} > 0$, but the inequality still holds trivially when $\prod_{i=1}^{n} Y_i^{a_i} = 0$. And so, by monotonicity,

$$\int \prod_{i=1}^{n} Y_i^{a_i} d\mu \leqslant \int \sum_{i=1}^{n} a_i Y_i d\mu = \sum_{i=1}^{n} a_i \int Y_i d\mu = 1$$

Which completes the proof.

□

Corollary 3.6.3.1 (Hölder's Inequality)**.** *If $p, q > 0$ s.t $1/p + 1/q = 1$, then for all measurable f and g,*

$$\int |fg| d\mu \leqslant \left(\int |f|^p d\mu \right)^{1/p} \left(\int |g|^q d\mu \right)^{1/q}$$

Proof. Let $X_1 = |f|^p$ and $X_2 = |g|^q$. □

Corollary 3.6.3.2 (The Cauchy-Schwarz Inequality)**.** *If f and g are measurable, then,*

$$\int |fg| d\mu \leqslant \left(\int f^2 d\mu \right)^{1/2} \left(\int g^2 d\mu \right)^{1/2}$$

Proof. Apply Hölder's Inequality with $p = q = 2$. □

3.7 Hilbert Spaces and $\mathcal{L}^p$ Spaces

The theory of Hilbert Spaces is quite general and far-reaching. A Hilbert space can be thought of as a generalization of Euclidean space, which includes the vector space $\mathbb{R}^n$ and the Euclidean dot product. More generally, a Hilbert Space is a vector space $\mathcal{H}$ with an inner product $\langle \cdot, \cdot \rangle$ which induces a complete metric space. Recall an inner product is a map $\mathcal{H} \times \mathcal{H} \to \mathbb{C}$ which satisfies,

- $\langle f, g \rangle = \overline{\langle g, f \rangle}$
- $\langle af_1 + bf_2, g \rangle = a\langle f_1, g \rangle + b\langle f_2, g \rangle$
- $\langle f, f \rangle \geqslant 0$

Where $\bar{\cdot}$ denotes complex conjugation. For our purposes though, we can ignore this distinction. If the inner product is a map to $\mathbb{R}$, we find it is symmetric and linear in both its arguments. The norm induced by the inner product is: $\|f\| = \sqrt{\langle f, f \rangle}$.

3.7.1 The $\mathcal{L}^p$ Spaces

Recall from our study of measurable functions that linear combinations of measurable functions are measurable. It is not hard to see from this that measurable functions constitute a vector space. In this space, we define the norm,

$$\|X\|_p = \left(\int X^p d\mu \right)^{1/p}$$

We show that this operation does induce a valid norm, and we construct the corresponding Hilbert space. We now prove the triangle inequality, which is a crucial step to verifying that we have a valid norm on our hands. A first step to realize is that, in the language of norms, Hölder's Inequality states that $|XY| \leqslant \|X\|_p \|Y\|_q$.

Theorem 3.7.1 (Minkowski's Inequality)**.** *The triangle equality holds in the norm* $\|\cdot\|_p$*:*

$$\|X + Y\|_p < \|X\|_p + \|Y\|_p$$

Proof. We proceed in the usual fashion. For now, assume that $0 < p < \infty$. The other cases are trivial. First, observe that $f(x) = x^p$ is convex for $x \geqslant 0$. And thus, for any x, y, $|x+y|^p = |\frac{1}{2}(2x) + \frac{1}{2}(2y)|^p \leqslant \frac{1}{2}|2x|^p + \frac{1}{2}|2y|^p = 2^{p-1}(|x| + |y|)$. Note the same will hold true using $1/p$.

$$\|X + Y\|_p = \left(\int |X + Y|^p d\mu \right)^{1/p} \leqslant 2^{\frac{p-1}{p}} \left(\int |X|^p + |Y|^p d\mu \right)^{1/p}$$

$$= 2^{1-\frac{1}{p}}\left(\|X\|_p^p + \|Y\|_p^p\right)^{1/p} \leqslant 2^{1-\frac{1}{p}}\left(\frac{1}{2}(2\|X\|_p^p)^{1/p} + \frac{1}{2}(2\|Y\|_p^p)^{1/p}\right)$$
$$= 2^{1-\frac{1}{p}}\left(2^{\frac{1}{p}-1}\|X\|_p + 2^{\frac{1}{p}-1}\|Y\|_p\right) = \|X\|_p + \|Y\|_p$$

To now address the $p \in \{0, \infty\}$ cases, recall that,

$$\|X\|_0 = \mu(\{|X| \geqslant 0\}) \qquad \|X\|_\infty = \text{ess-}\sup |X|$$

Where the ess-sup is defined as the smallest supremum of all functions Y equal to X almost everywhere:

$$\text{ess-}\sup X = \inf_{Y=X\ a.e.} \sup Y$$

So then, by subadditivity,

$$\|X+Y\|_0 = \mu(\{|X| \geqslant 0\} \cup \{|Y| \geqslant 0\})$$
$$\leqslant \mu(\{|X| \geqslant 0\}) + \mu(\{|Y| \geqslant 0\}) = \|X\|_0 + \|Y\|_0$$

Now for $p = \infty$. Intuiviely, the idea is just that $\max(X+Y) \leqslant \max(X) + \max(Y)$, but the almost-everywhereness adds some subtleties. Indeed, note that for any $X' = X$ a.e. and $Y' = Y$ almost everywhere, $X' + Y' = X + Y$ almost everywhere. Furthermore, $\sup\{X' + Y'\} \leqslant \sup\{X'\} + \sup\{Y'\}$. So then,

$$\|X+Y\|_\infty = \inf_{Z=X+Y\ a.e.} \sup|Z| \leqslant \inf_{X'=X, Y'=Y\ a.e.} |X' + Y'|$$
$$\leqslant \inf_{X'=X\ a.e.} \sup|X| + \inf_{Y'=Y a.e.} \sup|Y| = \|X\|_\infty + \|Y\|_\infty$$

This proves the claim for all p. □

we define the space $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$ as $\mathcal{L}^p(\Omega, \mathcal{F}, \mu) = \{X \in \mathcal{M}(\Omega, \mathcal{F}) : \|X\|_p < \infty\}$. Note it is possible for $\|X-Y\|_p = 0$ but for $X \neq Y$ everywhere. However, it is necessary and sufficient that $X = Y$ *almost everywhere*. And thus, we think of $\|\cdot\|_p$ as operating on the equivalence classes of functions, with equivalence if two functions are equivalent almost everywhere. Thus, $\|\cdot\|_p$ is not truly a norm (rather it is a pseudonorm) since $\|X - Y\| = 0$ does not imply $X = Y$. To make it a norm, we let L^p be the equivalence classes of $\mathcal{L}^p$.

Theorem 3.7.2. *$\|X - Y\|_p = 0$ if and only if $X = Y$ almost everywhere.*

Proof. Let $A = \{X \neq Y\}$. Now let $A_0 = \{|X - Y| \geqslant 1/2\}$ and $A_n = \{1/2^n > |X - Y| \geqslant 1/2^{n-1}\}$. It is clear to see that $A = \bigcup_{i=0}^n A_n$ and also that all the A_n's are disjoint. It therefore follows from countable additivity that $\mu(A) = \sum_{i=1}^n \mu(A_n)$. Then if If $\mu(A) > 0$, then there must be some $n^\star$ for which $\mu(A_{n^\star}) > 0$. Assume for now that $n^\star > 0$. It follows that,

$$\|X - Y\|_p^p = \int_{A_{n^\star}} |X-Y|^p d\mu + \int_{A^c_{n^\star}} |X-Y|^p d\mu \geqslant \frac{1}{(2^{n^\star - 1})^p} \mu(A_{n^\star})$$

(Here, we use the notation that $\int_C X d\mu = \int X\mathbb{I}\{C\}d\mu$). The above expression is obviously nonzero. If $n^\star = 0$, a similar analysis holds. Thus, it must be that $\mu(A) = 0$. This shows the "only if". For the "if", we have,

$$\|X - Y\|_p^p = \int |X - Y|^p d\mu = \int_A |X - Y|^p d\mu + \int_{A^c} |X - Y|^p d\mu$$

$\int_A |X - Y|^p d\mu = 0$ because A is negligible. $\int_{A^c} |X - Y|^p d\mu = 0$ because $X = Y$ everywhere in A^c. Thus the above is entirely nonzero. □

From the above, it's clear that $\|\cdot\|_p$ is a valid norm acting on L^p. In order to verify that L^p is a Hilbert space, it is necessary to verify that $\|\cdot\|_p$ induces a complete metric.

Theorem 3.7.3. *L^p is complete.*

Proof. Recall a space is complete if Cauchy sequences converge to a limit function. So let $X_1, X_2...$ be a Cauchy sequence. Here's a first step. Let n_k be a sequence of numbers such that for $n, m \geqslant n_k$, $\|X_n - X_m\|_p < 1/2^k$. It then follows that $\{X_{n_k}\}_k$ is a Cauchy sequence and $\sum_{k=1}^\infty \|X_{n_{k+1}} - X_{n_k}\|_p \leqslant 1$. Now, define $Y_k = \inf_{m \leqslant k} X_{n_m}$.

Proposition 14. $\liminf_k X_{n_k} = \limsup_k X_{n_k}$

Proof. Define $L_{i,j} = \inf_{i \leqslant k \leqslant j} X_{n_k}$. First observe that,

$$\|X_{n_i} - L_{i,j}\|_p = \left\| (X_{n_i} - Y_{i,j}) \sum_{k=i}^{j} \mathbb{I}\{Y_{i,j} = X_{n_k}\} \right\|_p$$

$$\leqslant \sum_{m=i}^{k} \left\| (X_{n_i} - Y_{i,j})\mathbb{I}\{Y_{i,j} = X_{n_k}\} \right\|_p \leqslant \sum_{m=i}^{k} \|X_{n_i} - X_{n_k}\|_p < \sum_{k=i}^{j} \frac{1}{2^i} \leqslant 2^{i-1}$$

Now, taking $j \to \infty$, we find $Y_{i,j} \to \inf_{k>i} X_{n_k}$ and thus $\|X_{n_i} - \inf_{k>iX_{n_k}}\|_p \leqslant 2^{i-1}$. A similar argument yields $\|X_{n_i} - \sup_{k>iX_{n_k}}\|_p \leqslant\leqslant \frac{1}{2^{i-1}}$. It follows that,

$$\|\liminf_{k \geqslant i} X_{n_k} - \limsup_{k \geqslant i} X_{n_k}\|_p \leqslant \|\inf_{k \geqslant i} X_{n_k} - \sup_{k \geqslant i} X_{n_k}\|_p$$

$$\leqslant \|\inf_{k \geqslant i} X_{n_k} - X_{n_i}\|_p + \|\sup_{k \geqslant i} X_{n_k} - X_{n_i}\|_p \leqslant 2\left(\frac{1}{2^{i-1}}\right) = \frac{1}{2^{i-2}}$$

Observe that, for all i, $\liminf_{k \geqslant i} X_{n_k} = \liminf_{k \geqslant 1} X_{n_k}$ and likewise for the $\limsup$. It thus follows that $\|\limsup_k X_{n_k} - \liminf_k X_{n_k}\|_p < \frac{1}{2^{i-2}}$ for all i, and thus, $\limsup_k X_{n_k} = \liminf_k X_{n_k}$ almost everywhere. □

It follows that X_{n_k} converges to a limit; call it X_∞. We will now show that $X_n \to X_\infty$ as well. This follows as a simple consequence of the triangle inequality and the definition of a Cauchy sequence. Indeed, take M such that $\|X_n - X_m\|_p < \epsilon/2$ for all $n, m \geqslant M$. Now let K be such that $k \geqslant K$ implies that $\|X_{n_k} - X_\infty\|_p < \epsilon/2$. Letting $N = \max(N, n_K)$, it follows that for all $n \geqslant N$,

$$\|X_n - X_\infty\|_p \leqslant \|X_n - X_{n_K}\|_p + \|X_{n_K} - X_\infty\|_p \leqslant \epsilon/2 + \epsilon/2 = \epsilon$$

Taking $\epsilon \to 0$, we achieve the desired result.

□

3.8 Convergence Notions

The last topic in this chapter is what it means, exactly, for one random variable to converge to another. We have just seen one: convergence in L^p. There are two more we must consider: *convergence in probability* and *almost sure convergence.*

Definition 17 (Convergence in Probability)**.** *We say $X_n \to X$ in probability if, for all $\epsilon > 0$, $\mathbb{P}(\{|X_n - X| > \epsilon\}) \to 0$*

Definition 18 (Almost Sure Convergence)**.** *We say $X_n \to X$ almost surely if $\mathbb{P}(\{\lim_n X_n = X\}) = 1$*

Definition 19 (Convergence in L^1)**.** *We say $X_n \to X$ in L^1 if $\|X_n - X\|_1 \to 0$*

Proposition 15. *Almost sure convergence implies convergence in probability*

Proof. We first show the statement about convergence in probability. Fix $\epsilon > 0$. First, we know that $\lim_n X_n$ exists almost surely. Simply write,

$$\begin{aligned}
1 = \mathbb{P}(\lim_n X_n = X) &= \mathbb{P}(\{\lim_n |X_n - X| = 0\}) \\
\leqslant \mathbb{P}(\{\lim_n |X_n - X| < \epsilon\}) &\leqslant \mathbb{P}(\lim_n \{|X_n - X| < \epsilon\}) \\
&= \lim_n \mathbb{P}(\{|X_n - X| < \epsilon\})
\end{aligned}$$

Where the last line is due to dominated convergence. □

Theorem 3.8.1. *Convergence in probability implies convergence in L^1.*

Proof. Suppose $X_n \to X$ in probability. Fix $\epsilon, \delta > 0$. Let N be such that $n \geqslant N \implies \mathbb{P}(\{|X_n - X| > \epsilon\}) < \delta$. Now observe that for $n \geqslant N$,

$$\begin{aligned}
\mathbb{P}(|X_n - X|) &= \mathbb{P}(|X_n - X|\{|X_n - X| > \epsilon\} + |X_n - X|\{|X_n - X| \leqslant \epsilon\}) \\
&\leqslant \epsilon + \mathbb{P}(|X_n - X|\{|X_n - X| > \epsilon\})
\end{aligned}$$

□