Introduction to Measure Theoretic Probability

Sam Leone

January 10, 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:" $\sigma(\mathcal{E}) = \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \Omega\}$. 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.*

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 1. *$\mathcal{I}$ 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 2. *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 2 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 3. *$\mathcal{I}$ 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 4. *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'_is. 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.

□

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 5. *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. □

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 5 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 6. *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 7. *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 8. *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. □

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. □

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 9. *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 $\square$

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 $\square$

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

Proof. Check the properties of a λ system (not hard) $\square$

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,

$$\begin{aligned} \mu^\star(A^c) &= \mu^\star(\Omega) - \mu^\star(A) \\ &= \mu'(\Omega) - \mu'(A) = \mu'(A^c) \end{aligned}$$

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 $\mu_i^\star(A) = \mu^\star(A \cap B_i), \mu_i'(A) = \mu'(A \cap B_i)$ for al $A \in \mathcal{B}(\mathbb{R}^d)$. It follows that,

$$\mu' = \sum_i \mu_i' \qquad \mu^\star = \sum_i \mu_i^\star$$

And since each $\mu_i', \mu_i^\star$ 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!