Measure Theory

Andreas Tsantilas
Professor Shatah

Spring, 2021

Contents

Introduction

Intro here

1 The Concept of Measure

1.1 σ-algebras

The idea of a σ-algebra is a very important concept to introduce the idea of a measurable space. [WRITE MORE]

Definition 1.1.1 (Topology)**.** A collection τ of subsets of a set X, or $\tau \subseteq \mathcal{P}(X)$, is known as a *topology* in X if the following conditions are satisfied:

(1) $\emptyset, X \in \tau$.

(2) If $V_i \in \tau$ for $i = 1, \dots, n$, then $V_1 \cap \cdots \cap V_n \in \tau$.

(3) If $\alpha \in \Lambda$, where Λ is an index set, then

$$\bigcup_{\alpha \in \Lambda} V_\alpha \in \tau.$$

Note that this applies to finite, countably infinite, or uncountably infinite unions.

If τ is a topology in X, then elements of τ are called *open sets* in X.

Definition 1.1.2 (σ-algebra)**.** A σ-algebra on a set X is a collection $\mathcal{M}$ of subsets of X, or $\mathcal{M} \subseteq \mathcal{P}(X)$, such that the following conditions hold:

(1) $\emptyset \in \mathcal{M}$.

(2) If $E \in \mathcal{M}$, then the complement $E^c := X \setminus E$ of E is also in $\mathcal{M}$. Combined with condition (1), we have that $X \in \mathcal{M}$.

(3) If $E_1, E_2, \dots \in \mathcal{M}$, then the countable union

$$\bigcup_{i=1}^{\infty} E_i \in \mathcal{M}.$$

A possible interpretation of $\mathcal{M}$ is the set of measurable subsets of X.

Definition 1.1.3 ($\mathcal{M}$-measurable Space)**.** The pair $(X, \mathcal{M})$, where X is a set and $\mathcal{M}$ is a σ-algebra, is known as a *$\mathcal{M}$-measurable space.*

Our idea of a measurable space makes sense; we ought to be able to assign a measure to each element in $\mathcal{M}$; we must include the empty set and thus its complement, X. Moreover, If we know the measure of a set, then we also want to know the measeure of its complement, and therefore have that the complements must also be in $\mathcal{M}$. Lastly, we want to ensure closure under countable unions. By DeMorgan's laws, we see that this implies closure under countable intersection as well. The stipulation of (3), we essentially are saying that if we can construct a set $A \subseteq X$ by taking a countable union, then there ought to be a well-defined measure assigned to it, even if it is infinite.

Indicated by the name, a σ-algebra has an algebraic structure, in its closure under the two operations of union and intersection. However, these σ-algebras themselves also form a σ-algebra under intersection.

Lemma 1.1.1 (Intersection of σ-algebras)**.** Let $\mathcal{M}_\alpha$ be a (possibly uncountable) collections of σ-algebras on X, and $\alpha \in \Lambda$. Then

$$\bigcap_{\alpha \in \Lambda} \mathcal{M}_\alpha$$

is also a σ-algebra.

Proof.

1. Clearly, $\emptyset \in \mathcal{M}_\alpha$ for all α.
2. Morever, if $E \in \cap\mathcal{M}_\alpha$, then $E \in \mathcal{M}_\alpha$ for all $\alpha \Rightarrow E^c \in \mathcal{M}_\alpha\ \forall\alpha \Rightarrow E^c \in \cap\mathcal{M}_\alpha$.
3. Lastly, suppose $E_1, E_2, \cdots \in \cap\mathcal{M}_\alpha$. Then $E_1, E_2, \cdots \in \mathcal{M}_\alpha\ \forall\alpha \Rightarrow \cup_{i=1}^\infty E_i \in \mathcal{M}_\alpha\ \forall\alpha \Rightarrow \cup_{i=1}^\infty E_i \in \cap\mathcal{M}_\alpha$.

□

Analogous to how groups can be generated, we can also define a procedure to generate σ-algebras.

Definition 1.1.4 (Generation of σ-algebras)**.** Let $\mathcal{F}$ be a subset of $\mathcal{P}(X)$. Then we define the σ-algebra generated by $\mathcal{F}$ to be the intersection of all σ-algebras that contain $\mathcal{F}$. We denote this construction as $\langle\mathcal{F}\rangle$.

Similar to generating groups from sets, this is the smallest σ-algebra which contains $\mathcal{F}$. That is, if $\mathcal{F} \subseteq \mathcal{M}$, then $\langle\mathcal{F}\rangle \subseteq \mathcal{M}$.

Lemma 1.1.2. The set $\langle\mathcal{F}\rangle$ is the smallest σ-algebra containing $\mathcal{F}$.

Proof. We already proved that $\langle\mathcal{F}\rangle$ is a σ-algebra, which clearly contains $\mathcal{F}$. Now we show that it is the "smallest" in the above sense. Note that the existence of such a smallest set is proved by construction.

Let $\mathcal{M}$ be the smallest σ-algebra containing $\mathcal{F}$. Then we need to show that $\langle\mathcal{F}\rangle = \mathcal{M}$. Clearly, $\langle\mathcal{F}\rangle$ contains $\mathcal{F}$, so $\mathcal{M} \subseteq \langle\mathcal{F}\rangle$. Conversely, we know that $\mathcal{M}$ is a σ-algebra containing $\mathcal{F}$. Then $\langle\mathcal{F}\rangle \subseteq \mathcal{M}$ since $\mathcal{M}$ is included in the intersections. Therefore, $\mathcal{M} = \langle\mathcal{F}\rangle$.

Moreover, this construction is unique. □

This leads us to the most important σ-algebra.

Definition 1.1.5 (Borel σ-algebra)**.** Let (X, τ) be a topological space. A *Borel σ-algebra* is the algebra generated by the open sets of X, or τ. That is,

$$\mathcal{B}[X] := \langle\tau\rangle.$$

Thus, given a set X, a Borel σ-algebra will contain the open sets of X, the closed sets of X, as well as the countable unions of closed sets (known as F_σ sets), and the countable intersections of open sets (known as G_σ sets), the countable intersections of F_σ sets, and so on.

If the topology is clear, we can also create this desctiption with a metric space (X, d). All we really need is the concept of an open set. The Borel σ-algebra in a sense contains the topology τ and encapsulates in a natural manner the sets we wish to measure.

1.2 Measures

Definition 1.2.1 (Measure)**.** Let $(X, \mathcal{M})$ be a measurable space. A map

$$\mu : \mathcal{M} \to [0, \infty]$$

is called a *measure* if the following conditions are satisfied:

(1) $\mu(\emptyset) = 0$.

(2) Let $\{E_k\}_{k=1}^\infty$ be a sequence of sets, and $E_i \cap E_j = \emptyset$ if $i \neq j$. Then

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

The first principle seems very reasonable; since the empty set will always be a part of any σ-algebra, we ought to define its generalized volume to be 0. What condition (2) gives us is a way to combine volumes in a reasonable manner. Moreover, since the infinite union of sets is still in the σ-algebra, then this is the most natural condition to impose on any set expressed as a countable union of sets.

Definition 1.2.2 (Measure Space)**.** Let X be a set, $\mathcal{M}$ be a σ-algebra, and μ be a measure. Then the tuple

$$(X, \mathcal{M}, \mu)$$

is known as a *measure space.*

Example. Let X be a set, and $\mathcal{M} = \mathcal{P}(X)$. For a set $A \in \mathcal{M}$, we can define the counting measure

$$\mu(A) := \begin{cases} \#A & \#A < \infty \\ \infty & \#A = \infty \end{cases}.$$

Clearly, this satisfies the properties of a measure, since the empty set has zero elements, and for finite unions, the measure of the union is the sum of the measures. For infinite unions, then this μ clearly gives us ∞.

Example (Dirac Measure)**.** We call the Dirac measure for an element $p \in X$ to be

$$\delta_p(A) := \begin{cases} 1 & p \in A \\ 0 & p \notin A \end{cases}$$

where $A \in \mathcal{M}$. A way to think of this is like a point charge in physics, where the total charge enclosed in a surface depends only on whether or not the charge is contained in the surface.

Note how this idea is distinct from a *metric space*. Next, we wil consider some key properties of measures.

Theorem 1.2.1. Let $(X, \mathcal{M}, \mu)$ be a measure space. Then the following are true:

(1) (Monotonicity.) If $E_1, E_2 \in \mathcal{M}$ and $E_1 \subseteq E_2$, then $\mu(E_1) \leq \mu(E_2)$.

(2) (Countable Subadditivity.) If $E_1, E_2, \dots \in \mathcal{M}$, then

$$\mu\left(\bigcup_{n=0}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu(E_n).$$

(3) (Continuity from below). Suppose we have an increasing, exhaustive sequence of measurable sets, given by $E_1 \subseteq E_2 \subseteq E_3 \subseteq \dots$ and such that $\cup E_n = E$. Then

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

(4) (Continuity from above). Suppose we have the decreasing sequence such that $E_1 \supseteq E_2 \supseteq \dots$ and $\cap E_n = E$ and $\mu(E_1) < \infty$. Then

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

Definition 1.2.3 (Finite Measure)**.** A measure μ is said to be *finite* if

$$\exists E_1, E_2, \dots, \bigcup_{n=1}^{\infty} E_n = E : \ \mu(E_n) < \infty, \ \forall n \in \mathbb{N}.$$

1.3 The Lebesgue Measure

Our aim has been to construct a measure satisfying some nice properties in $\mathbb{R}^d$; those of satisfying intuitive notions of "volume" and translation invariance, in addition to the properties of a measure. Formally statesd, we want to satisfy:

(1) The measure of a box characterized by a vector $x = (x_1, x_2, \ldots, x_d)$ is $\mu(x) = \prod_{i=1}^{d} x_i$, and

(2) the measure of a set E is invariant under translations: $\mu(v + E) = \mu(E)$ for any vector $v \in \mathbb{R}^d$.

Ideally, we would like to construct a measure on all of $\mathcal{P}(\mathbb{R})$ satisfying (1) and (2). However, we will now show that a meaningful measure cannot be constructed on the entire power set.

Theorem 1.3.1. There is no measure μ satisfying properties (1) and (2) on all of $\mathcal{P}(\mathbb{R})$.

Proof. Let μ be a measure such that $\mu([0,1]) < \infty$ and satisfying property (2). We will prove that the only such measure is the trivial measure (i.e., $\mu = 0$), so that property (1) won't be satisfied, completing the proof.

We can partition the reals into cosets of $\mathbb{Q}$. We create the quotient group $\mathbb{R}/\mathbb{Q} = \{x + \mathbb{Q} : x \in \mathbb{R}\}$. This means that the cosets partition $\mathbb{R}$, and each is dense in $[0,1]$, so each coset [FILL IN DETAILS] $\square$

1.4 Measurable Functions

Definition 1.4.1 (Measurable Map)**.** Let $(\Omega_1, \mathcal{M}_1)$ and $(\Omega_2, \mathcal{M}_2)$ be measurable spaces. A function

$$f : \Omega_1 \to \Omega_2$$

is measurable with respect to $\mathcal{M}_1, \mathcal{M}_2$ if $f^{-1}(E) \in \mathcal{M}_1$ for all $E \in \mathcal{M}_2$.

If we recall our motivations for developing a theory of measures, we can see why such a formulation might be useful. Consider a function that is 1 on the interval $[0,1]$, and 0 otherwise. Then we want to find the set E that is sent to 1 under f; that is, $f^{-1}(\{1\})$. Then the integral will simply be $\mu(f^{-1}(\{1\}))$, since $\mu(E)$ can be thought of as the "volume." Therefore, the preimage $f^{-1}(\{1\})$ must be in the sigma algebra.

Example. Let $(\Omega, \mathcal{M})$ be a measurable space and let $(\mathbb{R}, \mathcal{B}[\mathbb{R}])$ be another measurable space. We can define an indicator function $\chi_E : \Omega \to \mathbb{R}$ for an arbitrary set $E \in \mathcal{M}$:

$$\chi_E(\omega) = \begin{cases} 1 & \omega \in E \\ 0 & \omega \notin E \end{cases}$$

For all measurable E, this is a measurable map; since if we look at the preimages, $\chi_E^{-1}(\emptyset) = \emptyset$, $\chi_E^{-1}(\mathbb{R}) = \Omega$, $\chi_E^{-1}(\{1\}) = E$, and $\chi_E^{-1}(\{0\}) = E^c$. However, since $E \in \mathcal{M}$, then we know that $E^c \in \mathcal{M}$. Therefore χ_E is a measurable function for measurable E.

Theorem 1.4.1 (Composition of Measurable Maps)**.** Let $f : (\Omega_1, \mathcal{M}_1) \to (\Omega_2, \mathcal{M}_2)$ and $g : (\Omega_2, \mathcal{M}_2) \to (\Omega_3, \mathcal{M}_3)$ be measurable functions. Then the composition

$$(g \circ f) : \Omega_1 \to \Omega_3$$

is $\mathcal{M}_1, \mathcal{M}_3$-measurable.

Proof. We have that $(h \circ f)^{-1}(E) = f^{-1}(h^{-1}(E))$. Since h is a measurable function, $h^{-1}(E) \in \mathcal{M}_2$. Similarly, since f is a measurable function, then $f^{-1}(h^{-1}(E)) \in \mathcal{M}_1$ for all $E \in \mathcal{M}_3$. $\square$

2 Integration

2.1 Integrating Nonnegative Functions

In order to construct the integral, we must first define the notion of partitioning the set into measurable parts.

Definition 2.1.1 ($\mathcal{M}$-partition)**.** Let $\mathcal{M}$ be a σ-algebra on X. An $\mathcal{M}$-partition is a finite colllection of disjoint elements A_i of $\mathcal{M}$ such that $A_1 \cup \cdots \cup A_n = X$.

The way we will build the integral is by defining the integral for a special class of functions, known as "simple functions." These are very nearly like step functions and should be reminiscent of upper and lower Riemann sums.

Definition 2.1.2. Let $(X, \mathcal{M})$ be a measurable space. Let $A_1, \ldots, A_n \in \mathcal{M}$. If $f : X \to \mathbb{R}$ is a measurable funcition expressible as

$$f(x) = \sum_{i=1}^{n} a_i \cdot \chi_{A_i}(x)$$

then f is refered to as a *simple function*. This representation may not be unique.

For instance, the Dirichlet function is a simple function, expressed as $1 \cdot \chi_{\mathbb{Q}}(x) + 0 \cdot \chi_{(R \setminus \mathbb{Q})}(x)$. Note that simple functions differ from step functions in that the A_i don't have to be unbroken intervals $[a, b]$; they need only be in the σ-algebra.

Definition 2.1.3 (Lower Lebesgue Sum)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space. Let $f : X \to [0, \infty]$ be an $\mathcal{M}$-measurable function and let P be the $\mathcal{M}$-partition $A_1, \ldots, A_n$. Then we define the *lower Lebesgue sum* to be

$$\mathcal{L}(f, P) = \sum_{i=1}^{n} \mu(A_i) \inf_{A_i} f.$$

As we get a finer and finer partition, the value of $\mathcal{L}$ only increases, analogous to how the lower Riemann sum increases monotonically with each refinement. Therefore, we can now define the integral:

Definition 2.1.4 (Integral of a Nonnegative function)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space. Let $f : X \to [0, \infty]$ be a Nonnegative $\mathcal{M}$-measurable function. Then the *integral of f with respect to μ* is

$$\int f d\mu = \sup\{\mathcal{L}(f, P) : P \text{ is an } \mathcal{M}\text{-partition of } X\}.$$

2.2 Convergence Theorems

Theorem 2.2.1. Suppose that $f_n \to f$ almost everywhere, and $|f_n(x)| \leq M$ almost everywhere, and $\mu(E) < \infty$. Then

$$\lim_{n\to\infty} \int_E f_n = \int_E \lim_{n\to\infty} f_n.$$

Proof. The equation is true (trivial) if the convergence is uniform, since

$$\int_E |f - f_n| \leq \sup_E |f - f_n| \mu(E)$$

and since the measure is finite and the supremum goes to 0, then this is true. By Egorov's theorem, we know that convergence is almost uniform, except on a set of arbitrarily small measure. Consider

$$\int_E (f_n - f) = \int_G (f_n - f) + \int_B (f_n - f)$$

The first goes to 0, and the second term is $2M\mu(B)$. Then this goes to zero since it is arbitrarily small. $\square$

Recall the "enemies" from before. These are functions which are unbounded on small sets, sets such that the domain has infinite measure. Suppose we are given a fixed positive function $g \geq 0$, and the integral of g is finite. Then we can make two observations:

(i) For all $\varepsilon > 0$, $\exists \delta$ such that

$$\mu(F) < \delta \Rightarrow \int_F g < \varepsilon.$$

so g controls (1), since things can go wrong only on small sets.

(ii) For all $\varepsilon > 0$, $\exists R > 0$ such that

$$\int_{B_R^c(0)} g < \varepsilon$$

where $B_R^c(0)$ is the complement of the open ball centered around a point.

Suppose we take any function that goes all the way up to infinity, like $1/x^2$. The only redeeming quality is that it is integrable :

$$\int g < \infty.$$

After restricting it, we have a bounded function on a bounded set. This is because since for any integrable function we are approaching it from below, then we can get as close to it as we want, but with a bounded function which is supported on a compact set.

Theorem 2.2.2 (Fatou's Lemma)**.** Let $f_n \geq 0$. Then

$$\int \liminf f_n \leq \liminf \int f_n$$

Supose that $f_n \to f$ almost everywhere. Then the $\liminf f_n = \lim f_n$. Then we have that

$$\int \lim f_n \leq \liminf \int f_n$$

Suppose $f_n \geq 0$ and f_n approaches f from below (i.e., monotonically). If we approach things in a monotone fashion, then

$$\int \lim f_n = \lim \int f_n$$

which trivially follows from Fatou's lemma.

Proof. We have that $f_n(x) \leq f(x)$. Thus

$$\limsup \int f_n(x) \leq \int f(x).$$

Thus from Fatou's lemma, then

$$\int f(x) \leq \liminf \int f_n(x).$$

Then we conclude that

$$\limsup \int f_n(x) \leq \int f(x) \leq \liminf \int f_n(x)$$

From which we get the desired result. □

Theorem 2.2.3 (Dominated Convergence Theorem)**.** Suppose $f_n(x)$ be such that $f_n \to f$ almost evertwhere, and let g be a function such that

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

and

$$\int g(x) < \infty.$$

Then

$$\lim \int f_n = \int \lim f_n.$$

2.3 Modes of Convergence

These different types of convergence are here in order to make a more complete picture.

Having covered the space L^1, we can now talk about the space L^p.

2.4 L^p Spaces

We will first start with the set $L\infty$, since it is defined differently. We would like this space to be the set of all bounded functions; however, we know that there may be sets of measure zero on which the function may be unbounded. Then we want to define the *essential supremum* of a function to be

$$\text{ess sup}\, f := \inf \left\{ a : \mu(f^{-1}(a, \infty)) = 0 \right\}.$$

Then over a domain $\mathcal{D}$, we define the space L^∞,

$$\|f\|_\infty = \text{ess sup}_{x \in \mathcal{D}} |f(x)|$$

We now need to prove that this is indeed a norm.

1. Clearly, $\|f\|_\infty = 0 \iff f = 0$ a.e.,
2. Moreover, by the triangle inequality, $\|f+g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.
3. Lastly, we know that we can multiply by scalar constants.

However, from the criteria of the $\|\cdot\|_\infty$, we can say that this norm is analogous to uniform convergence.

Consider the sequence of functions $f_n = g_n$ almost everywhere, where $|g_n| < M$. We can do this to bound the set everywhere on a set of measure zero. These g_n's can be thought of as being representatives on the class of functions. Then because of this bounded condition, $g_n \to g$ uniformly for $x \in \mathcal{D}$. Then we have that

$$\sup_{x \in \mathcal{D}} |g_n - g| \to 0$$

where we don't need to use the essential supremum since we define g to be 0 whenever f is unbounded, and equal to f otherwise.

2.5 L^p for $p \geq 1$

We can define the norm on $L^p(E)$ to be

$$\|f\|_p = \left(\int |f|^p fx \right)^{1/p}.$$

The only thing that is not yet clear is why is this a norm? It is a norm by a simple fact, elucidated by the following problem. Consider

$$f(x) = ax - \frac{x^p}{p}$$

This function has a maximum value:

$$f(x) \leq \frac{a^{p'}}{p'}, \ \frac{1}{p'} + \frac{1}{p} = 1.$$

Therefore, we get the inequality

$$|ax| \leq \frac{x^p}{p} + \frac{a^{p'}}{p'}$$

This is a generalization of the arithmetic and geometric mean, where $p = 2$.

Given $f, g \in L^p$, we can consider the product of the numbers

$$\frac{f(x)}{\|f\|_p} \frac{|g(x)|}{\|g\|_p} \leq \frac{1}{p} \left(\frac{|f(x)|}{\|f\|_p} \right)^p + \frac{1}{p'} \left(\frac{|g(x)|}{\|g\|_p} \right)^{p'}$$

If we integrate these functions, we get the inequality

$$\frac{1}{\|f\|_p \|g\|_p} \int |f(x)||g(x)| \leq 1$$

Which implies the following general case of

Theorem 2.5.1 (Hölder)**.** For $f, g \in L^p$,

$$\int |f(x)||g(x)| \leq \|f\|_p \|g\|_p .$$

We want to show that this is a norm.

Proof.

1. If $\|f\|_p = 0$, then $f = 0$ almost everywhere by the properties of integrals.
2. Next, we also have that $\|\lambda f\|_p = |\lambda| \, \|f\| \, |$.
3. Lastly, we just need to show the triangle inequality. We have that

$$\begin{aligned} \int |f+g|^p &= \int |f+g|^{p-1}|f+g| \\ &\leq \int |f+g|^{p-1} (|f| + |g|) \\ &\leq \int |f+g|^{p-1}|f| + \int |f+g|^{p-1}|g| \end{aligned}$$

Then from Hölder's inequality, this question is the same as asking if $f \in L^p$, then is $|f+g|$ in p'. We have that $p' = p/(p-1)$. We can raise the sum to the power p' to get $|f+g|^p$. Then if we apply Hölder to those terms,

$$\int |f+g|^p \leq \|f\|_p \|f+g\|_p^{p-1} + \|f+g\|_p^{p-1} \|g\|_p$$

Then we have

$$\|f+g\|_p^p \leq \|f+g\|_p^{p-1} \left(\|f\|_p + \|g\|_p \right).$$

From which we get the desired result.

□

Is there a relation between functions for different p's? Asuume that $\mu(\mathcal{D}) = \infty$. What does it mean for a function to be in $f \in L^1$? It means that it has a finite integral.

From Hölder's inequality, if the measure of the domain is finite, then

$$\|f\|_1 \leq \mu(\mathcal{D})^{1/p'} \|f\|_p$$

then we have that

$$L^p \subset L^q, \; q \leq p.$$

Theorem 2.5.2 (Riesz)**.** G

2.6 Product Measure and Fubini's Theorem

The principal goal of Fubini's theorem is to be able to t

Given two spaces X and Y, we can construct a measure on the set $X \times Y$. Ideally, given measure spaces $(X, \mathcal{M}_1, \mu)$ and $(Y, \mathcal{M}_2, \nu)$, we want to construct a measure space $(X \times Y, \mathcal{M}, \lambda)$ with the measure $\lambda = \mu \times \nu$. If you recall the construction of the measure on the real line, we began with the semialgebra, and then extended the measure from that onto the whole space by showing that the measure was countably additive. Therefore,

Definition 2.6.1 (σ-finite)**.** A measure space $(X, \mathcal{M}, \mu)$ is said to be σ-finite if X can be expressed as the countable union of sets of finite measure.

If we consider the semialgebra for X and Y respectively, we worked with half-open intervals of the form $(a, b]$. Thus, in our new set $X \times Y$, the semialgebra of choice is the set of rectangles R, of the from $(a, b] \times (c, d]$. This is indeed a semialgebra, since

1. The intersection of two rectangles R_1 and R_2 is also a rectangle, and
2. The complement of a rectangle is the countable union of rectangles.

The only thing left to show is that we have countable additivity of the measure λ.

Claim. The *pre-measure* λ is countably additive on the set of rectangles. That is, if $A \times B \in \mathcal{R}$, and $A \times B = \cup_{i \geq 1} A_i \times B_i$, then

$$\lambda(A \times B) = \sum_{i \geq 1} \lambda(A_i \times B_i).$$

Proof. If the hypothesis holds, then we have that

$$B = \cup_{\{i : x \in A_i\}} B_i$$

and so $\nu(B) = \sum \nu(B_i)$. Then $\nu(B)\chi_A(x) = \sum_{i\geq 1} \nu(B_i)\chi_{A_i}(x)$ since the function $\chi_A(x)$ can be expressed as a sum of disjoint subsets of A. Therefore, we can take the integral:

$$\int \nu(B)\chi_{A_i}(x)\ d\mu = \int \sum_{i\geq 1} \nu(B)\chi_{A_i}(x)\ d\mu = \sum_{i\geq 1} \int \nu(B)\chi_{A_i}(x)\ d\mu$$

which holds by the MCT. Therefore, if we integrate with respect to the measure μ,

$$\lambda(A\times B) = \sum_{i\geq 1} \mu(A_i)\nu(B_i) = \sum_{i\geq 1} \lambda(A_i\times B_i)$$

as desired. $\square$

Then suppose $E \in R_{\sigma\delta}$, which is the set of all countable intersection and unions of rectangles. Then we define the slices of E to be:

$$\begin{aligned} E^y &= \{x \in X : (x,y) \in E\} \subset X \\ E^x &= \{y \in Y : (x,y) \in E\} \subset Y. \end{aligned}$$

The slice E^x corresponds to holding x fixed, and sweeping along the set Y. This is so, since if $E = \cup_i R_i \in R_\sigma$, then $\chi_{E^x}(y) = \sup_i \chi_{R_i}(x,y) = \sup_i \chi_{R_i^x}(y)$ is ν-measurable. This is because the supremum for measurable sets is always measurable. Similarly, $F \in \cap_i E_i \in R_{\sigma\delta}$, then $\chi_{F^x}(y) = \inf_i \chi_{E_i}(x,y) = \inf_i \chi_{E_i^x}(y)$ which is measurable by a similar argument.

We may prefer to think of slices as some infinitesimal over which we can integrate. Indeed, for $E \in R_\sigma$, we have $E = \cup_{i\geq 1} A_i \times B_i$ and $\nu(E^x) = \sum_{i\geq 1} \nu(B_i)\chi_{A_i}(x)$. Then

$$\int \sum_{i\geq 1} \nu(B_i)\chi_{A_i}(x) d\mu = \sum_{i\geq 1} \nu(B_i)\mu(A_i) = \lambda(E)$$

since λ is countably additive. Thus, what we have shown is that we can do iterated integrals on an R_σ set.

Next, we can consider $R_{\sigma\delta}$ sets. We can iterate integrals on these by using the dominated convergence theorem. If we are taking intersections of sets (recall this is what the "δ" means), then we are dominated by the first set, or the largest one. We consider our set $E \in R_{\sigma\delta}$, given by the decreasing sequence $E = \lim_i E_i$, $E_i \in R_\sigma$, and $E_i \supseteq E_{i+1}$.

Now if the measure $\lambda(E) < \infty$, we can select the largest set in the decreasing sequence such that $\lambda(E_1) < \infty$. Then immediately, our integral is dominated by this. But since it is measurable,

$$\int \nu(E^x) d\mu = \lim_i \int \nu(E_i^x) d\mu = \lim_i \lambda(E_i) = \lambda(E)$$

since each E is an R_σ set, and the dominated convergence theorem.

Now take any measurable set E. Then E can be written as a disjoint union $F \cup Z$, where $\lambda(Z) = 0$, $\lambda(F) = \lambda(E)$, and $F \in R_{\sigma\delta}$. Then we can write

$$\int \nu(E^x) d\mu = \int \nu(F^x) + \nu(Z^x) d\mu = \int \nu(F) d\mu = \lambda(F) = \lambda(E).$$

Now that we have proved this for any measurable set, we know that this ability to integrate holds for simple functions which are compactly supported:

$$\int \sum_i^N \alpha_i \nu(E_i) d\mu = \sum_i^N \int \alpha_i \nu(E_i^x) d\mu = \sum_i^N \alpha_i \lambda(E_i).$$

Invoking littlewood's principles, any measurable function can be approximated by functions whose support is a set of finite measure. Here is where we need the definition of σ-finiteness. If μ and ν are σ-finite, then we can extend our result for simple functions to positive measurable functions, by approximating these functions by increasing sequences of simple functions which are zero outside of a set of finite measure. Then because of the monotone convergence theorem, if

$$f = \lim_{N\to\infty} \sum a_{i,N}\lambda(E_{i,N})$$

then

$$\lim_{N\to\infty} \sum_i^N \int \alpha_{i,N}\chi_{E_i}(x)d\mu = \int \lim_{N\to\infty} \sum_i^N \alpha_i \nu(E_i^x)d\mu = \int f^x d\mu.$$

Then we can state the result of Fubini:

Theorem 2.6.1 (Fubini)**.** Let (Ω_1, X, μ) and (Ω_2, Y, ν) be σ-finite measure spaces, and let $(\Omega_1 \times \Omega_2, X \times Y, \lambda)$ where $\lambda := \mu \times \nu$. Then if if $f \in L^1$ with respect to the product measure λ,

$$\int_X \left(\int_Y f d\nu \right) d\mu = \int_Y \left(\int_X f d\mu \right) d\nu.$$

There is a difficulty in the above theorem; it may be difficult to verify that $f \in L^1$ with respect to the product measure. However, there is a way around it; we can take f to be a positive function, and so we do not get any cancellations. Therefore, infinite integrals are allowed. That is the essence of Tonelli's Theorem:

Theorem 2.6.2 (Tonelli)**.** Suppose $f \geq 0$. Then

$$\int_X \left(\int_Y f d\nu \right) d\mu = \int_Y \left(\int_X f d\mu \right) d\nu.$$

Note that the integrals may be infinite.

If f is not positive, we can check if

$$\int_X \left(\int_Y |f(x,y)| d\nu \right) d\mu < \infty.$$

Then if this holds, we know we can use Fubini's Theorem.

Now suppose we are given a positive function $f : X \to \mathbb{R}$. If we look at the integral

$$\int f(x)d\mu(x) = \int \mu\left(\{x : f(x) > t\}\right) dt.$$

However, this result comes out of Fubini's theorem:

$$\int f(x)d\mu(x) = \int \left(\int_0^{f(x)} dt \right) d\mu(x) = \int \int H\left(f(x) - t\right) dt d\mu$$

Where the Heaviside function is

$$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Then we can write

$$\int \left(\int H(f(x) - t) d\mu(x) \right) dt$$

Where the Heaviside function can now be expressed as a characteristic function

$$\int \int \chi_{\{f(x)>t\}}(x) d\mu(x) dt = \int \mu\left(\{x : f(x) > t\}\right) dt.$$

If we take ϕ to be an increasing, C^1 function, where $\phi(0) = 0$, then we can verify that

$$\int \phi\left(f(x)\right) d\mu(x) = \int \mu\left(\{x : f(x) > t\}\right) d\phi(t)$$

so the integral becomes a Stieltjes intgral with ϕ. This follows from Fubini and the realization that

$$\phi(f(x)) = \int_0^{f(x)} \phi'(s) ds.$$

3 Probability

Everything we have said about measure theory can be neatly translated to probability by altering some key concepts slightly. However, probability starts and analysis ends when we talk about independence. This notion is the key to understanding the difference.

Our probability space is the measure space (Ω, Σ, P), where $P(\Omega) = 1$. We refer to sets $A \in \Sigma$ as events, and we say that two events are independent if

$$P(A \cap B) = P(A)P(B)$$

We define a random variable to be a measurable function:

$$X : \Omega \to \mathbb{R}.$$

The set Ω can in practice be extremely complicated. However, this map X induces a measure on the σ-algebra:

$$X : (\Omega, \Sigma, P) \to (\mathbb{R}, \mathcal{B})$$

Where some sets are

$$\{\omega \in \Omega : X \in (a, b]\} = F(b) - F(a).$$

Then if we want to integrate X, we can write

$$\int X(\omega) dP(\omega) = \int x dF$$

which comes from the fact that we can integrate by parts, since everything is finite.

If we are given two random variables, X, Y, then

$$X : (\Omega, \Sigma, P) \to (\mathbb{R}, \mathcal{B}, \mu) \ \mu = dF$$
$$Y : (\Omega, \Sigma, P) \to (\mathbb{R}, \mathcal{B}, \nu) \ \nu = dG$$

IF X and Y are independent, we mean that if we take the measure of two inverse images, then they are inependent. This means that the joint probability distribution is in fact a product measure. Since everything is finite, then Fubini holds.

Suppose we want to compute the following probability:

$$P(\{\omega : X + Y \leq t\}) = \iint_{x+y \leq t} d\mu d\nu.$$

This becomes

$$\int F(t - y) d\nu(y) = \int F(t - y) dG(y)$$

If G is differentiable, then we may prefer to think of this as a convolution:

$$\int F(t - y) G'(y) dy.$$

Therefore, the distribution function of two independent variables is just a convolution.

3.1 Independent Random Variables

All we need for the codomain of the random varible X is a set and a σ algebra. The measure induced on $\mathcal{B}$ by the random variable (measurable function) X is

$$F(x) = P\left(\{\omega \in \Omega : X(\omega) \leq x\}\right).$$

We could have written from the very beginning that

$$\mu_X(B) = P\left(\{X^{-1}(B)\}\right)$$

however, we only really need it for intervals. Note the distinction between X and x; X is a function $X(\omega)$, and x is a real number.

3.1.1 Inedependent Events

Recall that an event is just an element of the σ-algbra. Two events are independent if

$$P(A \cap B) = P(A)P(B).$$

We can say that these A, B are independent if the measures generated by the random variables X, Y if they are independent. Then

$$P\left(\omega : X(\omega) \in A \text{ and } Y(\omega) \in B\right) = P(\omega : X(\omega) \in A)P(\omega : Y(\omega) \in A).$$

Then the measure in the x, y plane is a product measure if the measure λ on the whole plane is $\mu_X \times \mu_Y$.

3.2 Law of Large Numbers

There are two ways of stating the law of large numbers.

If we have a sequecne of random variables $\{X_i\}_{i=1}^{\infty}$ which have finite mean and variance, where the mean is

$$\int_{\Omega} X(\omega) dP(\omega) = \int x d\mu_X.$$

The variance of x is assumed to be finite, where

$$Var(X) = \int_{\Omega} \left|X - \overline{\mu}\right|^2 dP =< \infty.$$

This automatically tells us that

$$\int X^2 dP < \infty$$

or that $X \in L^2(\Omega)$. Then this means that X is definitely in L^1, since everything is finite. This is the same as saying that if a random variable has a second moment, it has a first moment. Its very often to use the random variable

$$Y = X - \overline{\mu}$$

so that the mean of Y is 0.

If we have these, we want to look at there sum. This is nothing more than a sum of functions. However, beyond convergence theorems, there is not much we can say about this sum. One question we want to ask is how much does this sum differ from its mean? Then we get that

$$\int \sum_{i=1}^{N} X_i - N\overline{\mu}\ dP = 0$$

We should take the X_i to be pairwise independent. Then we can change

$$X_i \mapsto X_i - \overline{\mu}$$

Therefore these are independent random variables with mean 0. Thus,

$$\int X_i X_j dP = \int X_i dP \int X_j dP = 0$$

but since each mean is 0, this means that X_i and X_j are orthogonal in some way, with respect to the L^2 inner product. Then we can show that

$$\int X_1^2 X_2 dP = \int X_1^2 dP int X_2 dP.$$

We know that if X_1 and X_2 are independent, than any function of the two are independent. Then consider $F(X)$ and $G(Y)$. Then

$$\begin{aligned}
\int F(X)G(Y)dP &= \int F(x)G(y)d\lambda \\
&= \iint F(x)G(y)d\mu_X d\nu_Y \\
&= \int F(X)d\mu_X \int G(Y)d\mu(y)
\end{aligned}$$

Now let's think of the example of flipping a coin. Let's say that the coin is fair. Then we can consider $X(H) = 1$ and $X(T) = -1$. Then this is a function with mean 0. Then after N flips, we have N independent variables, all with mean 0. These all have a variance. Then all of them will have the same variance, since

$$Var(X_i) = \sigma^2.$$

Now we want to ask if the following converges:

$$F_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Or in what sense does $F_n \to F$? Recall that the best we can hope for is L^∞ convergence. Since everything is of finite measure, almost everywhere is equivalent to almost uniform.

Recall that the interpretation of pointwise convergence in a measure-theoretic sense is the same as saying

$$B_n = \{|F_n - F| > \delta\}$$
$$\limsup B_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} B_k$$

Convergence in measure means $\mu(B_n) \to 0$.

3.2.1 Sum of Independent Random Variables

The sum of N independent random variables with mean 0 is

$$\sum_{i=1}^{N} X_i$$

has variance

$$\begin{aligned}\int\left(\sum_{i=1}^{N} X_i\right)^2 dP &= \int (X_1 + X_2 + \cdots + X_N)^2\, dP \\ &= \int X_1^2 dP + \int X_2^2 dP + \cdots + \int X_N^2 dP \\ &= N\sigma^2.\end{aligned}$$

Then because of independence, things do not grow as quickly. In particular,

$$\int\left(\frac{\sum_{i=1}^{N} X_i}{N}\right) = \frac{\sigma^2}{N}$$

as $N \to \infty$, this goes to 0. We therefore have convergence in L^2, which implies convergence in measure, via Chebyshev's inequality.

Theorem 3.2.1 (Weak Law of Large Numbers)**.** Suppose $X_1, \ldots, X_n$ are independent random variables. Then

$$P\left(\left\{\left|\frac{1}{N}\sum_{i=1}^{N} - \overline{\mu}\right| > \delta\right\}\right) \to 0.$$

Can we change this from weakly going to 0 into goint to 0 almost everywhere? More specifically, which conditions can we add? Recall that convergence in measure means that

$$P(B_n) \to 0$$

and almost everywhere means that

$$\lim_{k\to\infty} P\left(\bigcup_{n=k}^{\infty} B_n\right)$$

In order to convert the first into the second, recognize that convergence in measure is the same as a sequence of numbers, b_n, going to 0. However, since measure is subadditive, then we can require that

$$\sum_{n=k}^{\infty} b_n = 0 \iff \sum_{n=1}^{\infty} b_n < \infty$$

which fulfills the condition of convergence almost everywhere. This gives us the following lemma:

Lemma 3.2.2 (Borel-Cantelli)**.** If $\sum P(B_n) < \infty$, then

$$P\left(\limsup B_n\right) = 0$$

The inverse is true with the added assumption of independence.

Under the condition that the mean is 0, and the variance is finite, then we got that $b_n \to 0$. We proved this because we gained an N from the variance. Let's take the integral of the next even number power:

$$\begin{aligned}\int\left(\sum X_i\right) dP &= \sum X_i^4 + \sum X_i^2 X_j^2 \\ &= N\int X_i^4 + 3(N)(N-1)\left(\int X_i^2\right)^2\end{aligned}$$

Now if we divide our original by N^4, we gain an N^2 in the denominator, which makes this summable. Now let us formally prove this with Chebyshev. Then Now we have that

$$F_N = \frac{1}{N}\sum_{i=1}^{N} X_i$$

then by Chebyshev,

$$P\left(|F_N| > \delta\right) \leq \frac{1}{\delta}\int |F_N|^4 dP$$

and so the above is

$$P(B_n) \leq \frac{C}{\delta^4 n^2}$$

which is summable. Therefore, we have almost everywhere convergence. Now we can get a rough set of conditions where we can get the Strong Law of Large nubmers. In reality, the Strong Law can be proven under a much weaker set of conditions.

For instance, for the weak law, we can do away with the requirement that the variance be finite. In order to use the L^1 norm, we can use Chebyshev to conclude that

$$P(F_N) \leq \frac{1}{\alpha}\int_{\{x : F_N(x) > \alpha\}} |F_N| dP.$$

If we can get the right hand side to go to zero, then we can prove the weak law. However, this distinction falls between the difference between L^1 and L^2. In particular, on a set of finite measure,

$$\int |f| dm \leq \sqrt{\int |f|^2 dm}.$$

The main difference only occurs for large values, or rather in how they approach infinity. Thus, we can chop our function X_i at level K, in order to bound its value. Then

$$X_i^K = \begin{cases} X_i & |X_i| \leq K \\ 0 & |X_i| > K. \end{cases}$$

Since this is a bounded function on a set of finite measure, it is now in L^2 (in fact, it is in every L^p space). Then we consider the remainder, or what we chopped off:

$$Y_i^K = \begin{cases} 0 & |X_i| \leq K \\ X_i & |X_i| > K. \end{cases}$$

Then we can write

$$P(\{|Y_i^K| > \alpha\}) \leq \frac{1}{\alpha}\int_{K < |Y_1|} |Y_1|$$

where we only need to consider Y_1 since all are identically distributed. If a function is integrable in the L^1 sense, then its integral goes to 0 as $K \to \infty$. Then we can re-write our function

$$F_N = \frac{1}{N}\sum X_N^K + \frac{1}{N}\sum Y_N^K = \xi_N^K + \eta_N^K.$$

We can show that

$$\{|F_N| > \alpha\} \subset \{|\xi_n^k| > \alpha/2\} \cup \{|\eta_n^k| > \alpha/2\}$$

Where we used the fact that if $a+b>c$, then either a or b must be greater than $c/2$. Then

$$P\left(\{|F_n|>\alpha\}\right) \leq P\left(\{|\xi_n^k|>\alpha\}\right) + P\left(\{|\eta_n^k|>\alpha\}\right).$$

Since ξ is L^2, the first term goes to 0. Moreover, the right term is independent of n, since it is bounded by

$$\frac{2}{\alpha}\int_{K<|Y_1|}|Y_1|dP$$

However, as K gets larger, we can make this arbitrarily small. Therefore for all ε,

$$P\left(\{|F_n|>\alpha\}\right)<\varepsilon \Rightarrow P\left(\{|F_n|>\alpha\}\right)=0.$$

If X_i are i.i.d. random variables, we have that the central limit theorem tells us that we see a Gaussian distribution if we zoom our scale out from not $1/n$, where we see concentration around the mean, but $1/\sqrt{n}$.

How can we relate measures to things we know? Suppose $f\geq 0$, $f\in L^1(\mu)$. We define Local L^p convergence to mean

$$f\in L^1_{loc} \iff \int_{B_r(0)}|f|<\infty.$$

If we create a measure

$$\nu(A)=\int_A f d\mu$$

this is a new measure on $(\mathbb{R}^d,\mathcal{B},\nu)$. If this is the case, we write

$$\frac{d\nu}{d\mu}=f.$$

This has the important quality that of $\mu(E)=0$, then $\nu(E)=0$.

Definition 3.2.1 (Absolute Continuity)**.** Let μ,ν be measures. Then we say that ν is absolutely continuous with respect to μ if

$$\mu(E)=0\Rightarrow\nu(E)=0.$$

This is denoted by

$$\nu\ll\mu.$$

This statement is equivalent to $\forall\varepsilon>0$, $\exists\delta$ such that if

$$\mu(E)<\delta\Rightarrow\nu(E)<\varepsilon.$$

Theorem 3.2.3. Let $f\in L^1(\mu)$. Then $\forall\varepsilon>0$, $\exists\delta$ such that if $\mu(E)<\delta$, then

$$\int_E|f|\;d\mu<\varepsilon.$$

Proof. We can prove this using the technique of truncating the function at large values, and intersecting the domain with a ball. Therefore, the error we make in both directions is small. ☐

We can then immediately conclude that if ν can be expressed as the integral of some integrable function f with respect to μ, then $\nu\ll\mu$.

3.3 Fundamental Theorem of Calculus

Recall from a first calculus course the following fundamental theorem:

Theorem 3.3.1 (Fundamental Theorem of Calculus)**.** Let $F \in C^1$ and $f \in C^0$. Then

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

(2)
$$\int_a^x F'(t)dt = F(x) - F(a)$$

The proof is not hard for $f \in C^0$ and $F \in C^1$, which can be proved by calculus methods. The question is, is this still true for worse functions?

Let $G(x)$ be equal to $\int_a^x f(t)dt$. It is easy to show that G is continuous, since

$$|G(x) - G(x_0)| = \left|\int_{x_0}^x f(t)dt\right| \leq \int_{x_0}^x |f(t)|dt < \varepsilon$$

for sufficiently small $|x_0 - x|$. Therefore, this is a continuous function in x, as long as $f \in L^1$, which follows from theorem [THEOREM].

We might ask if G is differentiable, which is the same as asking if we can differentiate an integral. The answer is yes, almost everywhere. Then we have the two questions:

(1) Which functions can we differentiate almost everywhere a.e.?

(2) Whic functions are the integral of their derivative?

The latter point is often taken for granted, but not necessarily true. When we exclude sets of measure 0; consider the familiar Heaviside function, whose derivative is 0 almost everywhere. Worse still, the Cantor-Lebesgue function is monotone between 0 and 1, but its derivative is 0 almost everywhere.

We proved functions which are integrals are continuous. They are also absolutely continuous, like the measure. If F is monotone (Bounded Variation), then F is differentiable almost everywhere. Moreover, we will prove that if $f \in L^1$, then

$$\frac{d}{dx}\int_a^x f(t)dt = f(x).$$

Lastly, if f is absolutely continuous, then

$$\int_a^x F'(t)dt = F(x) - F(a).$$

3.4 The Radon-Nikodym Theorem

We learned that if $f \in L^1(d\mu)$, and $f \geq 0$, then we know that the following defines a measure on the set E:

$$\nu(E) = \int_A f d\mu(x)$$

which is a measure, since it is countably additive by the monotone convergence theorem. If we get rid of the positivity requirement for f, then we still have a measure which satisfies $\nu(\varnothing) = 0$ and is countably additive.

Definition 3.4.1. A *signed measure* is a measure μ such that

1. $\mu(\varnothing) = 0$
2. μ is countably additive.

In this section, we want to explore the question: given a measure μ belonging to $(\mathbb{R}, \mathcal{B}, \mu)$, when can we write that

$$\mu((a, b]) = \int_a^b f(x) dm(x)?$$

Example. Let the measure μ be the dirac measure,

$$\mu(A) = \begin{cases} 1 & 0 \in A \\ 0 & 0 \notin A \end{cases}.$$

If this were true, we could use the dominated convergence theorem to show that

$$0 = \int_{\{0\}} f(x) dx = \lim_{n\to\infty} \int_{B_{1/n}(0)} fx dx = 1$$

which is a contradiction.

Now for the sake of argument, assume that

$$\nu(E) = \int_E f d\mu.$$

We can say a few things about this; in particular, we observe that this implies that $\mu(A) = 0 \Rightarrow \nu(A) = 0$. This is the same as absolute continuity. We see tha this condition can help avoid a situation as in above. Surprisingly, this turns out to be the only condition we need. If this condition is true, we can write ν as an integral.

Before we can show the Radon-Nikodym theorem, we must understand how to differentiate. We take the difference of two functions over a distance, and send the distance to 0. This may no longer be positive, hence the introduction of a signed measure.

If $f \in L^1$, it means that

$$\int |f| = \int f^+ + \int f^- < \infty.$$

Thus we can do this for every L^1 function. If μ is a signed measure, can we do something like

$$\mu = \mu^+ - \mu^-?$$

μ^+ is positive on every subset of our space, and μ^- is also positive on every subset of the space. Such a splitting may not be unique, but they differ on a set of measure 0.

Let μ be a signed measure which we wish to split. Keep in mind the following picture. Then

Definition 3.4.2 (Total Positivity)**.** A set A is totally positive if

$$\mu(B) \geq 0$$

forall $B \subseteq A$.

Given a set A, and $\mu(A) > 0$, then does A have a totally positive subset? The answer is yes, we can do this iteratively. Suppose A has a negative set. We can take

$$\inf_{E \subset A} \mu(E) < 0$$

which must be less than 0. Then we know that $\exists E_1$ such that

$$\mu(E_1) < \frac{1}{2} \inf_{E \subset A} \inf \mu(E)$$

if we have a sequence $\mu(E_n) \to \inf$. E_1 may not be a totally negative set. All we know is that

$$\mu(A \setminus E_1) \geq \mu(A).$$

Now check if $A \setminus E_1$ has a negative subset. If not, we're done. If so, we can take a set E_2 such that

$$\mu(E_2) < \frac{1}{2} \inf_{E \subset A \setminus E_1} \inf \mu(E).$$

We know that the measure of $A \setminus \cup E_n$ is increasing. For simplicity, if $\mu(A)$ is finite. Then we assume that

$$\sup_{B \in \mathcal{B}} |\mu(B)| < \infty.$$

Because of the finiteness,

$$\sum |\mu(E_n)| < \infty.$$

This tells us that $|\mu(E_n)| \to 0$. Therefore,

$$\mu(E_{n+1}) < \frac{1}{2} \inf_{E \in A \setminus \cup E_n} \mu(E)$$

where E is what makes the measure negative. Then since the E_n goes to 0, we know that the quantity on the right must be greater than or equal to 0. Therefore,

$$A \setminus \bigcup_{n=1}^{\infty} E_n.$$

Then we let the positive part of a space X as

$$X^+ = \bigcup_{\alpha} A_\alpha.$$

We can repeat the similar processes for X^-. Now we have two positive measures, defined on different sets:

$$\mu(A) = \mu^+(A) - \mu^-(A).$$

Since these are disjointly supported, we can write $\mu^+ \perp \mu^-$. That is,

$$supp\mu^+ \cap supp\mu^- = Z$$

where $\mu(Z) = 0$.

3.4.1 The Theorem

Theorem 3.4.1. If $\nu \ll \mu$, then there exists a function $f \geq 0$ such that

$$\nu(E) = \int_E f d\mu.$$

Proof. We proved the converse before. Now suppose that $\nu \ll \mu$. □

If we want to find an f, we know that f must satisfy the following:

$$\int_A f d\mu \leq \nu(A).$$

Now consider all positive functions f such that

$$\int_A f d\mu \leq \nu(A).$$

We have equality when we consider

$$\sup\left\{\int_X f d\mu : f \geq 0,\ \int_A f d\mu \leq \nu(A)\right\}.$$

We aim to get equality from the monotone convergence theorem. Let this supremum be denoted by I^*, which is a number. We can find a sequence f_n such that $\int f_n \to I^*$. Since $f_n \geq 0$, we can modify it to be monotone by a slight change:

$$g_n(x) = \max(f_1(x), f_2(x), \ldots, f_n(x)).$$

Then g_n converges by the monotone convergence theorem, and f_n converges by the dominated convergence theorem. Therefore,

$$\int f(x) d\mu(x) = I^* \leq \nu(X).$$

Now consider

$$\nu(A) - \int_A f d\mu - \varepsilon\mu(A).$$

Since we are taking the difference of two measures, we have a signed measure. Now this signed measure cannot have a positive set. This will prove the theorem.

Assume $\exists Y^+ \subset X$ which is totally positive. If $A \subset Y^+$, then

$$\nu(A) - \int_A f d\mu - \varepsilon\mu(A) \geq 0.$$

However, if we consider the function $f + \varepsilon\chi_A$, then this will be part of the functions we try to maximize over. Then the only way we can have

$$\nu(A) \geq \int_A f + \varepsilon\chi_A$$

is if $\mu(A) = 0 \Rightarrow \nu(A) = 0$ by absolute continuity. Therefore, there is only a negative part to the above measure. Then we get that

$$\nu(A) - \int_A f d\mu - \varepsilon\mu(A) \leq 0$$

which means

$$0 \leq \nu(A) - \int_A f d\mu \leq \varepsilon\mu(A) \leq c\varepsilon$$

where the left side comes from the fact that $I^* \leq \nu(A)$. Therefore, we can let ε be arbitrarily small, which proves that

$$\nu(A) = \int_A f d\mu.$$

3.5 Derivatives

In the last section, we encountered the concept of the Radon-Nikodym derivative. If $\nu \ll \mu$, then there exists and $f \in L^1(d\mu)$ where

$$\nu(A) = \int_A d\mu = \int f\chi_A d\mu$$

where we can write

$$\frac{d\nu}{d\mu} = f.$$

Suppose we have a finite measure on the real line. For every finite measure, we can automatically assign a function, given by

$$F(x) = \mu((-\infty, x])$$

which is monotone, finite, right continuous. The measure μ of a set A is

$$\mu(A) = \int \chi_A d\mu = \int \chi_A dF$$

where "dF" is not a derivative, but denotes integration with respect to a measure. Thus $d\mu$ and dF are the same from this point of view. The aim of the Radon-Nikodym theorem is that any function F can be written as the sum of 3 functions:

$$F = F_{ac} + J + g$$

where

(1) f is absolutely continuous, which will give us a measure which is absolutely continuous with respect to the Lebesgue measure. In particular, if F is absolutely continuous, the next two terms are not present, which is what we have from Radon-Nikodym.

(2) J is a jump function, which accounts for the discontinuities

(3) g is a function which is continuous but whose derivative is 0 almost everywhere.

The function g is very strange, and such a function would be the Cantor-Lebesgue function, for example.

We want to know when our F satisfies the fundamental theorem of calculus. Suppose we are given a function $f \in L^1$. Then the following is pointwise differentiable almost everywhere, then we have

$$\frac{d}{dx}\left(\int_a^x f(t)dt\right) = f(x) \text{ a.e.}$$

We can understand a general proof. We have to show that

$$\frac{1}{h}\int_x^{x+h} f(t)dt \to f(x).$$

This is true if f is continuous; however, if f is not continuous, then we do not know if the limit exists. However, we know that the supremum over all h always exists. Then

Definition 3.5.1 (Maximal Function)**.** The *maximal function* in $\mathbb{R}^d$ is defined as

$$M(f)(x) := \sup_r \frac{1}{\mathrm{Vol}(B_r(x))}\int_{B_r(x)} |f(y)|dy$$

Where $B_r(x)$ is the d-dimensional open ball centered around x

We can transform the above into

$$\frac{1}{h}\int_x^{x+h} f(t)dt = \int_{-\infty}^{\infty} \frac{1}{h}\chi\left(\frac{x-t}{h}\right) f(t)dt$$

Which you should recognize as a convolution with the characteristic function $[-1,0]$. The limit of the convolution with a compactly supported function is always $f(x)$, since in the limit

$$\lim_{n\to\infty}\int n\phi(n(x-y))f(y)dy = f(x)$$

or approximate identity, when $\int\phi = 1$ and $\operatorname{supp}\phi$ is compact. Therefore, we have that

$$\lim_{h\to 0}\int_x^{x+h} f(t)dt = f(x)$$

in the L^1 sense. If something converges in L^1, it converges in measure. Then we know that convergence in measure implies a subsequence where we have almost-everywhere convergence.

Theorem 3.5.1. There exists a subsequence $h_n \to 0$ such that

$$\frac{1}{h_n}\int_x^{x+h_n} f(t)dt = f(x) \text{ a.e.}$$

However, we would like this to be independent of the subsequence we choose. This is where we can use the concept of the so-called "maximal function."

What does this function look like? It can be thought of describing the local behavior in terms of large x. Moreover,

$$m\left(\{x : M(f) > \alpha\}\right) \le \frac{c}{\alpha}\|f\|_1$$

3.6 Covering Lemmae

Take a collection of infinitely many closed balls in $\mathbb{R}^d$, $0 < R < \infty$. Then the claim is that we will be able to find disjoint balls B_n such that

$$\sum \text{Vol}(B_n) \ge \frac{1}{5^d}.$$

Let $\sup_B R = M$, and group all balls together such that $M/2 \le R \le M$. Then take the maximum number of such balls that do not intersect one another. Then any other ball in the group $M/2 \le R \le M$ which is not selected must touch at least one of the balls. If we multiply its radius by 5, it will encompass the ball in the group which it touches.

Next, consider balls whose radii are $M/4 \le R \le M/2$. Then we now have disjoint balls such that

$$\bigcup_{i=1}^{\infty} B_i \subseteq \bigcup_\alpha B_\alpha$$

And so after we have blown the balls up,

$$\bigcup 5B_i \supseteq \bigcup_\alpha B_\alpha$$

Then we have that

$$\sum_{i=1}^{\infty}|5B_i| \ge \left|\bigcup_\alpha B_\alpha\right| \Rightarrow \sum_{i=1}^{\infty}\text{Vol}(B_i) \ge \frac{1}{5^d}\text{Vol}\left(\bigcup_\alpha B_\alpha\right).$$

Now we have found $E \subset \cup_\alpha B_\alpha$, and we took a fractional bite out of our set. Now under which codnitions can we repeat this? Let K be very large so that $E \subset \cup^K B_i$. Then we can make a fine cover [Define]

3.7 Differentiation

4/5 We showed that if $f \in L^1$, then

$$F(x) = \int_0^x f(t)dt$$

is well-defined and an absolutely continuous function. Moreover, we proved that

$$\frac{1}{h}\int_x^{x+h} f(t)dt \to f(x)$$

almost everywhere.

Now, we want to ask which types of functions can be differnetiated, or for which functions does

$$\lim_{h\to 0} \frac{F(x+h) - F(x)}{h}$$

exist? It will turn out that if F is monotone, then this limit exists almost everywhere. The first thing we can consider for such a function F is the lim sup and lim inf. It makes a difference if h is positive or negative. Therefore,

[PROVE]

However, just because a function is differentiable almost everywhere, doesn't mean that the fundamental theorem of calculus applies. Again, consider the Heaviside function. What is true, however, that

$$\int_a^b F'(x)dx \leq F(b) - F(a).$$

This can be proven by Fatou's lemma.

When is the fundamental theorem satisfied? We have that F is monotone, and $F'(X) = f(x) \geq 0$ a.e. Then we have that this integral is bounded on $[a, b]$ since it is less than or equal to $F(b) - F(a)$. Then

$$\int_a^x f(t)dt = G(x)$$

is differentiable almost everywhere from our maximal function. $G'(x) = F'(x)$, except G has absolute continuity, which F does not have.

Theorem 3.7.1. If F is absolutely continuous (or of bounded variation), then

$$F(x) - F(a) = \int_a^x f(t)dt$$

If we take μ to be Lebesgue, then we have that

4 The Fourier Transform

4/5

In this section, we will define the Fourier transform and understand its basic qualities, as well as how it relates to the concept of measure.

Definition 4.0.1 (Fourier Transform of a function)**.** Let $f \in L^1(\mathbb{R})$. Then we define *Fourier Transform of* f, denoted $\mathcal{F}(f)$, to be

$$\mathcal{F}(f)(\xi) = \int_{\mathbb{R}} e^{-ix\xi} f(x)\, dx. \tag{1}$$

It should be clear that this integral exists, since the norm of $e^{ix\xi}$ is always 1. Therefore, such a concept is well-defined if $f \in L^1(\mathbb{R})$. Moreover, it is continuous in ξ, since we can dominate this integral by the DCT:

Lemma 4.0.1. The Fourier transform of an L^1 function f is continuous in ξ.

Proof. Let $(\xi)_n$ be a sequence converging to ξ^*. Foremost,

$$\int \left|e^{-ix\xi} f(x)\right|\, dx = \int \left|e^{-ix\xi}\right| |f(x)|\, dx = \int |f(x)|\, dx < \infty$$

since $\left|e^{-ix\xi}\right| = 1$, and $f \in L^1(\mathbb{R})$. Then, due to the dominated convergence theorem, we can exchange the limit and the integral.

$$\begin{aligned}
\lim_{\xi_n \to \xi^*} \mathcal{F}(f)(\xi_n) &= \lim_{\xi_n \to \xi^*} \int e^{-ix\xi_n} f(x)\, dx \\
&= \int \lim_{\xi_n \to \xi^*} e^{-ix\xi_n} f(x)\, dx \\
&= \int e^{-ix\xi^*} f(x)\, dx \\
&= \mathcal{F}(f)(\xi^*).
\end{aligned}$$

□

Moreover, it should be verified by the Riemann-Lebesgue lemma that

$$\mathcal{F}(f)(\xi) \to 0 \text{ as } \xi \to \infty.$$

Such a proof is more trivial if f is differentiable, and $f' \in L^1(\mathbb{R})$.

We will consider the following argument by Littlewood's principles. Let f be approximately a smooth, compactly-supported function. That is, f is "almost" $\phi \in C_0^\infty(\mathbb{R})$) in the sense that $\|f - \phi\|_{L^1} < \varepsilon$. Then

$$\left|\int e^{-ix\xi}(f(x) - \phi(x))\, dx\right|$$