Math 202ANotes

Kanyes Thaker

Fall 2020

Introduction

Intervals in the real line carry a very simple notion of size: their length. The Lebesgue theory of measure is a vast generalization of this notion. It assigns to a huge collection of subsets of the real line a notion of size which builds in a consistent fashion from the special case of intervals. It sets the framework for the theory of Lebesgue integration, which offers a much more powerful notion of the integral than Riemann's, the theory taught in introductory real analysis. This class will present these two theories rigorously, building them in the context of motivating problems in probability theory and analysis. We will also develop the theory of metric and topological spaces, including the study of function spaces, such as the set of continuous functions $f : \mathbb{C} \mapsto \mathbb{R}$ on a given compact set, or the L_p spaces, for which the Lebesgue theory will be needed. We will prove topological results such as the theorems of Tychonoff, Urysohn, Tietze,and function-space results including the Arzelà-Ascoli and Stone-Weierstrass theorems.

Formally the only prerequisite is Math 104. However, this is a graduate-level course, (although many students entering high-level math PhD programs encounter the material as undergraduates). Those whose primary preparation is Math 104, or an equivalent introductory course to real analysis, should expect significantly greater sophistication in the treatment offered here and more challenging problem sets than in the earlier class. Not all details will be treated in lectures, when comprehensiveness may sometimes be sacrificed for the sake of clarity. Students maybe expected to study texts including reading and analysing assigned portions.

Contents

1 Intro – A Brief Review

We have a fundamental notion of approximation and distance in which we have $\mathbb{Q}$ embedded as a dense set in $\mathbb{R}$. The rationals, of course, are those numbers of the form m/n where $m, n \in \mathbb{Z}$. The distance measure between two rationals is $d(r, q)$ is itself a positive rational by basic algebra (i.e. $d : \mathbb{Q} \times \mathbb{Q} \mapsto \mathbb{Q}_+$) and allows us to define a metric space along with the triangle inequality. Every rational also has a "recurring binary structure" ($1/2 = 0.500000...$, $1/9 = 0.11111....$), but most real numbers do not. Imagine a situation where we assign a binary number to every number between 0 and 1. We then know that there's a probability of 1/2 that our number is strictly less than 1/2 and a probability of 1/2 that our number is strictly greater than 1/2, but the probability that the number is itself 1/2 is 0. Intuitively this seems paradoxical; there is a 0% chance that we will manage to hit a "perfect" rational number, so we intuitively guess that while the rationals are dense they are also sparse in the reals.

1.1 Construction of the Reals

Suppose a sequence $(x_n \in \mathbb{Q} : n \in \mathbb{N})$. The sequence is Cauchy if $|x_n - x_m| \to 0$ as $n, m \to \infty$. More formally, we can define an epsilon-delta statement as $\forall \epsilon > 0 \exists N \in \mathbb{N} : \forall m, n \geq N,\ |x_n - x_m| < \epsilon$. Perhaps, then $\mathbb{R} \simeq$ a Cauchy sequence of rationals. Perhaps here we've overengineered what we're doing. If we identify two Cauchy sequences of rationals, x_n and y_n, we don't want to accidentally count a real number twice. We define a notion of equality – if $x_n - y_n \to 0$, we denote those sequences as equivalent. Recall that equivalence relations must follow reflexivity, symmetry, and transitivity, and that relations divide a space into equivalence classes. The real numbers are precisely the set of equivalence classes of the set of Cauchy sequences of the rationals.

1.2 Riemann Integration

The Riemann integral divides the domain of a function into intervals and takes an infimum and supremum approximation of rectangular area of the function. Then $f : [0, 1] \mapsto \mathbb{R}$ is Riemann integrable if $\exists I \in \mathbb{R}$ such that $\forall \epsilon > 0$, and we define $mesh(s)$ is the maximum interval size, then $\exists \delta > 0$ such that $\forall$ partitions s of $mesh(s) < \delta$, then $|I_s(f) - I| < \epsilon$.

2 Borel's Normal Number Theorem

It is difficult to develop a probabilistic interpretation of measure if we limit ourselves to discrete spaces. A complete interpretation of probability based in measure theory requires us to consider two types of problems – an infinitely repeated operation (infinite coin flips) and infinitely fine operations (selecting a single point from a segment).

2.1 The Unit Interval

We want to express both of the aforementioned problems in a single model. To this end, we recall the notions of **probabilistic independence** and **probabilistic expected value**. We form analogies here from general mathematical notions such as the interval and the Riemann integral.

Let Ω denote $(0, 1]$, with $\omega \in \Omega$. The **length** of the interval $I = (a, b]$ is denoted $|I|$, with

$$|I| = |(a, b]| = b - a. \tag{2.1}$$

For a finite collection of disjoint intervals I_i contained in Ω, the union of the intervals A has probability

$$P(A) = \sum_{i=1}^{n} |I_i|. \tag{2.2}$$

If A and B are two such finite disjoint unions of intervals, and A and B are themselves disjoint, then

$$P(A \cup B) = P(A) + P(B) \tag{2.3}$$

All this should be familiar from basic probability theory. What is less obvious is that this relation is a consequence of the linearity of the Riemann integral:

$$\int_0^1 (f(\omega) + g(\omega))d\omega = \int_0^1 f(\omega)d\omega + \int_0^1 g(\omega)d\omega \tag{2.4}$$

In this way, we can model the instant an "event" happens during a unit interval of time as random in that it lies in A with probability $P(A)$. Here we have our "point-in-a-segment" problem modeled. We can also use $P(A)$ to model our coin flips.

Associate with each ω the nonterminting dyadic expression

$$\omega = \sum_{n=1}^{\infty} \frac{d_n(\omega)}{2^n} = .d_1(\omega)d_2(\omega)... \tag{2.5}$$

with each $d_n(\omega)$ drawn from $\{0, 1\}$. Thus the sequence of $d_i(\omega)$'s is precisely the sequence of binary digits in the expansion of ω. Numbers with multiple binary expansions take the nonterminating expansion for definiteness. We can split the unit interval into different intervals corresponding to the binary expansion; $d_1(\omega) = 0$ implies that ω is in $(0, \frac{1}{2}]$, while $d_1(\omega) = 1$ puts ω in $(\frac{1}{2}, 1]$, and so on. Dividing the unit interval into these intervals, **dyadic intervals**, of width $\frac{1}{2^n}$ essentially allows us to model the "coin-flipping" problem within the unit interval.

2.2 The Weak Law of Large Numbers

The Weak Law of Large Numbers

For all $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left[\omega : \left|\frac{1}{n}\sum_{i=1}^{n} d_i(\omega) - \frac{1}{2}\right| \geq \varepsilon\right] = 0 \tag{2.6}$$

In essence, as n tends to infinity, the probability that the "relative frequency" of heads will deviate from 1/2 tends to 0.

2.2.1 Proof of the WLLN

Applying the Riemann integral in the role of the expected value (as mentioned before), we can prove (1.6) using Chebyshev's inequality. This becomes simpler through the use of the Rademacher functions:

Rademacher Functions

$$r_n(\omega) = 2d_n(\omega) - 1 = \begin{cases} +1 & \text{if } d_n(\omega) = 1 \\ -1 & \text{if } d_n(\omega) = 1 \end{cases} \tag{2.7}$$

We can then consider the partial sums

$$s_n(\omega) = \sum_{i=1}^{n} r_i(\omega). \tag{2.8}$$

Substituting $d_i(\omega)$ we see $\sum_{i=1}^{n} d_i(\omega) = (s_n(\omega)+n)/2$; we can then sustitute ε with $\varepsilon/2$ and rephrase (1.6) as

$$\lim_{n\to\infty} P\left[\omega : \left|\frac{1}{n}s_n(\omega)\right| \geq \varepsilon\right] = 0. \tag{2.8}$$

The Rademacher functions can be interpreted probabilistically – if you have a background in basic probability, you'll note the connection to the random walk on the real line, and that $r_i(\omega)$ is the distance moved at step i, and $s_i(\omega)$ is the position at step i. Note that every dyadic interval must contain two child dyadic intervals (don't think about this too hard), and therefore $r_i(\omega)$ has value $+1$ in one child and -1 in the other. By this logic, $\int_0^1 r_i(\omega)d\omega = 0$ and

$$\int_0^1 s_n(\omega)d\omega = 0, \tag{2.9}$$

here stating that the mean position after n steps is 0. Suppose $i < j$. Then for a dyadic interval of rank $j-1$, $r_i(\omega)$ is constant and $r_j(\omega)$ is -1 on the left and $+1$ on the right. Then

$$\int_0^1 r_i(\omega)r_j(\omega)d\omega = 0, \tag{2.10}$$

corresponding to the fact that independent random variables are uncorrelated. Since for each $r_i(\omega)$, $r_i^2(\omega) = 1$, we can additionally see that

$$\int_0^1 s_n^2(\omega)d\omega = n, \tag{2.11}$$

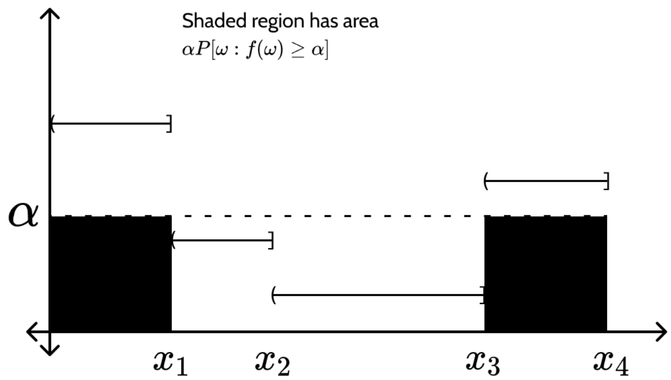
corresponding to the linearity of the variances of independent random variables.

Now we've made an extensive relationship between the Riemann integral of the Rademacher functions and probability theory. We can formally apply Chebyshev's inequality to (1.6) and get

$$P[\omega : |s_n(\omega)| \geq n\varepsilon] \leq \frac{1}{n^2\varepsilon^2}\int_0^1 s_n^2(\omega)d\omega = \frac{1}{n\varepsilon^2}. \tag{2.12}$$

To understand why we can do this, let's take a step back and investigate the forces at work. Let f be a step function $f(w) = c_j$ for $w \in (x_{j-1}, x_j]$ where $0 = x_0 < \ldots < x_k = 1$. Clearly $[\omega : f(w) \geq \alpha]$ is for $\alpha > 0$ a finite union of intervals. Then,

$$\alpha P[\omega : f(\omega) \geq \alpha] \leq \int_0^1 f(\omega)d\omega. \tag{2.13}$$



$\alpha = n^2\varepsilon^2$ and $f(\omega) = s_n^2(\omega)$ gives the desired result.

2.3 The Strong Law of Large Numbers

Normal Numbers

$$N = \left[\omega : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} d_i(\omega) = \frac{1}{2}\right] \tag{2.14}$$

(1.14) describes the **normal numbers**, i.e. the real numbers whose binary representation has a roughly "uniform" distribution of 0s and 1s. We attempt to show that it is practically certain that a number drawn at random from the unit interval is a normal number. Let $A \subset \Omega$ be **negligible** if for all $\varepsilon > 0$ there is a countable (but not necessarily disjoint) collection of intervals $\mathcal{I}_j$ such that

$$A \subset \bigcup_j \mathcal{I}_j, \quad \sum_k |\mathcal{I}_k| < \varepsilon. \tag{2.15}$$

A negligible set is one which can be covered by a set of intervals whose sum total length is arbitrarily small. Intuitively, $P(A) = 0$. As an observation, note that a countable union of negligible sets must be negligible (why?). A set containing a single point is negligible; since any countable sets is the union of countable "single point" sets, every countable set is negligible. The rationals are negligible – note that every rational number must be a repeating decimal in binary. If we treat the digits of the binary representation as a coin flip, there is 0 probability that a series of coin flips would repeat in a pattern infinitely.

Borel's Normal Number Theorem

The set of normal numbers has negligible complement.

The above theorem is a special case of the **strong law of large numbers**.

2.3.1 Proof of Borel's Normal Number Theorem

It is not enough to simply say that N^c is countable and therefore negligible, since N^c is uncountable (for instance take all numbers $0.11u_1 11u_2 11u_3...$ where the frequency of 1 is at least 2/3 and so is not normal; however, the cardinality of numbers of that form is the same as that of the uncountable set of numbers $0.u_1u_2u_3...$). Note that (1.14) is equivalent to saying

$$N = \left[\omega : \lim_{n\to\infty} \frac{1}{n} s_n(\omega) = 0\right] \tag{2.16}$$

For N^c to be negligible it must satisfy (1.15). Using the same logic used in the proof for the WLLN, we can write

$$P[w : |s_n(\omega)| \geq n\varepsilon] \leq \frac{1}{n^4\varepsilon^4} \int_0^1 s_n^4(\omega)d\omega. \tag{2.17}$$

We can take $s_n^4(\omega)$ to mean $\sum r_\alpha(\omega)r_\beta(\omega)r_\gamma(\omega)r_\delta(\omega)$.

Notice that some of these indices might be the same since they all range over $1, ..., n$ independently of each other. Using a different set of indices, we can see that our possible values of the summand in the above expression could be the following:

$$\begin{cases} r_i^4(\omega) = 1 \\ r_i^2(\omega)r_j^2(\omega) = 1 \\ r_i^2(\omega)r_j(\omega)r_k(\omega) = r_j(\omega)r_k(\omega) \\ r_i^3(\omega)r_j(\omega) = r_i(\omega)r_j(\omega) \\ r_i(\omega)r_j(\omega)r_k(\omega)r_l(\omega) \end{cases} \tag{2.18}$$

In the last 3 cases, note that they must necessarily integrate to 0. Consider, without loss of generality, that i is the greatest index in the 5th expression above. Note that on the $i-1$st dyadic interval, $r_j(\omega)$, $r_k(\omega)$, $r_l(\omega)$ are constant. Meanwhile, $r_i(\omega)$ is -1 on the left half of the interval and +1 on the right half, so the total integral is 0; the same is true for the 3rd and 4th expression. The first two expressions integrate to 1. In this case, how many times does each expression occur? There are n occurrences of case 1. There are $3n(n-1)$ occurrences of type 2 (n choices for index i, 3 other indices which can be paired with it, and $n-1$ choices for the remaining index). In that case:

$$\int_0^1 s_n^4(\omega)d\omega = n + 3n(n-1) \leq 3n^2. \tag{2.19}$$

Plugging this into (1.17) yields

$$P\left[\omega : \left|\frac{1}{n}s_n(\omega)\right| \geq \varepsilon\right] \leq \frac{3}{n^2\varepsilon^4}. \tag{2.20}$$

Let $\{\varepsilon_n\}$ be a positive sequence tending to 0 such that $\sum_n \varepsilon^{-4}n^{-2}$ converges. Then if $A_n = [\omega : |n^{-1}s_n(\omega)| \geq \varepsilon_n]$, then $\sum_n P(A_n) < \infty$ by (1.20). For some m, let ω lies in A_n^c for all $n \geq m$. Then $|n^{-1}s_n(\omega)| < \varepsilon_n$ for all $n \geq m$, then ω is normal by (1.16) because ε_n tends to 0. For each m $\cap_{n=m}^{\infty} A_n^c \subset N$, so $N^c \subset \cup_{n=m}^{\infty} A_n$.

Given ε we can choose m such that $\sum_{n=m}^{\infty} P(A_n) < \varepsilon$. A_n is a finite disjoint union of intervals $\cup_k I_{nk}$, with $\sum_k |I_{nk}| = P(A_n)$, so $\cup_{n=m}^{\infty} A_n$ is a countable union of intervals; the intervals I_{nk} provide the covering of N^c that the definition of negligibility calls for.

2.4 Strong Law vs Weak Law

Borel's normal number theorem is stronger than the weak law of large numbers. For all n let $f_n(\omega)$ be a step function on $(0, 1]$. Consider the following relations:

$$\lim_{n\to\infty} P[\omega : |f_n(\omega)| \geq \varepsilon] = 0 \tag{2.21}$$

$$\left[\omega : \lim_{n\to\infty} f_n(\omega) = 0\right]. \tag{2.22}$$

Notice that if we let $f_n(\omega) = \frac{1}{n}s_n(\omega)$, we get the weak law. (1.22) is then the definition of a normal number (1.14). We delay the proof for now, but there is a general result that shows that (1.22) implies (1.21) if (1.22) has negligible complement; however, the converse is not true, which means that the strong law implies the weak law but not the other way around.

2.5 Length

The complement of the set of normal numbers is negligible, so its complement cannot be; this would imply $(0,1] = N \cup N^c$ is negligible, which is clearly untrue. From this we get the (obvious) conclusion that an interval with positive length is nonnegligible. We then arrive at the following theorem which forms a basis for the Lebesgue theory of measure.

Interval Length Theorem

Let $I = (a, b]$ be an interval of length $|I| = b - a$. Consider a finite sequence of intervals $I_k = (a_k, b_k]$. These intervals need not be subintervals of $(0, 1]$.

1. If $\cup_k I_k \subset I$ and the I_k are disjoint, then $\sum_k |I_k| \leq |I|$.
2. If $I \subset \cup_k I_k$, where the I_k need not be disjoint, then $|I| \leq \sum_k |I_k|$.
3. If $I = \cup_k I_k$ and the I_k are disjoint, then $|I| = \sum_k |I_k|$.

These properties follow directly from the linearity of the Riemann integral.

3 Probability Measures

3.1 Spaces

We here introduce the notion of a space Ω with elements ω; in probability theory this represents the sum total outcome space for an experiment. Consider a subset of Ω to be an **event** and a point ω to be a **sample point**.

3.1.1 Assigning Probabilities

If we are treating Ω as a probability space, we need a method to assign probabilities to events. It would be incredible if we had some well-defined way to assign probabilities to events within a probabilistic space, no matter what that event may be. However, this is impossible. Instead, we work with subclasses of the class of all subsets of Ω – the **algebras** and **σ-algebras** (also called **fields** and **σ-fields**).

3.2 Classes of Sets

In order to continue with this special treatment of classes, we must determine that those classes are closed under set operations. As such, we need a class of sets that contains the intervals and is closed under countable union and intersection. Let's begin with the smallest possible example – the **singleton** $\{x\}$, which is a countable intersection $\cap_n(x - \frac{1}{n}, x]$ of intervals. Now if a class contains all the singletons and is closed under any arbitrary number of unions, then it *must* contain all subsets of Ω. This is clearly too extensive for us. Instead, we have to restrict ourselves to countable or finite set operations.

Algebra

A class $\mathscr{F}$ of subsets of Ω is an **algebra** if it contains Ω itself and is closed under complement and finite union. In other words:

1. $\Omega \in \mathscr{F}$.
2. $A \in \mathscr{F} \implies A^c \in \mathscr{F}$.
3. $A, B \in \mathscr{F} \implies A \cup B \in \mathscr{F}$.

By DeMorgan's law, $A \cap B = (A^c \cup B^c)^c$, and $A \cup B = (A^c \cap B^c)^c$, so if $\mathscr{F}$ is closed under complementation and finite union, it must also be closed under finite intersection. $\mathscr{F}$ is a **σ-algebra** if it is closed under *countable* unions and intersections (not just finite ones). A non-σ algebra is sometimes called **finitely additive**. A set in a class $\mathscr{F}$ is called an **$\mathscr{F}$-set**. Here we explore some examples of sets such sets.

1. Consider the set of finite disjoint subintervals of $\Omega = (0, 1]$. Augment this set with the empty set; the result is the algebra $\mathscr{B}_0$. If we let $A = (a_1, a_1'] \cup \ldots \cup (a_m, a_m']$, with each component interval disjoint, then $A^c = (0, a_1] \cup \ldots (a_m', 1]$ which is part of $\mathscr{B}_0$. Likewise, defining B analagously to A, $A \cap B \in \mathscr{B}_0$ and so $A \cup B \in \mathscr{B}_0$, therefore $\mathscr{B}_0$ is proven to be an algebra.
2. Consider the **finite-cofinite algebra**, i.e. $\mathscr{F}$ consisting of sets that are finite or whose complement is finite. Then $\mathscr{F}$ is an algebra and if Ω is finite, $\mathscr{F}$ is a σ-algebra.
3. Consider the **countable-cocountable algebra**, i.e. $\mathscr{F}$ consisting of sets that are countable or whose complement is countable. Then $\mathscr{F}$ is a σ-algebra. Additionally, note that a σ-algebra need not contain all

elements of Ω – in the case of the reals, there exist elements outside the algebra in Ω which can be constructed from uncountable unions.

The largest σ-algebra in Ω is the **power class** 2^Ω consisting of all subsets of Ω, while the smallest σ-algebra consists of Ω and $\emptyset$.

In a normal probability or analysis problem, we are concerned with a rather small or constrained class $\mathscr{A}$. The above examples show that it is easy to achieve sets outside of $\mathscr{A}$ using countable or finite operations. We want to consider sets that both contain $\mathscr{A}$ and are σ-algebras. To make things nicer, we awnt these sets to be as small as possible. To this end, we define the **σ-algebra generated by** $\mathscr{A}$ as the intersection of *all* algebras containing $\mathscr{A}$.

1. $\mathscr{A} \subset \sigma(\mathscr{A})$
2. $\sigma(\mathscr{A})$ is a σ-algebra
3. If $\mathscr{A} \subset \mathscr{G}$ and $\mathscr{G}$ is a σ-algebra then $\sigma(\mathscr{A}) \subset \mathscr{G}$.

Borel Sets and the Borel-σ-Algebra

Let $\mathscr{I}$ be the class of subintervals of $(0,1]$ and let $\mathscr{B} = \sigma(\mathscr{I})$. The elements of $\mathscr{B}$ are the **Borel sets** and $\mathscr{B}$ is the **Borel-σ-algebra**. From our previous examples, since $\mathscr{I} \subset \mathscr{B}_0 \subset \mathscr{B}$, $\sigma(\mathscr{B}_0) = \mathscr{B}$. $\mathscr{B}$ contains the normal numbers, as well as all open sets of $(0,1]$; in a way, it is the "biggest" σ-algebra.

3.3 Probability Measures

A **set function** is a real-valued function whose domain is a class of subsets of Ω. The set function P on an algebra $\mathscr{F}$ is a **probability measure** if:

1. $0 \le P(A) \le 1$ for $A \subset \mathscr{F}$
2. $P(\emptyset) = 0$ and $P(\Omega) = 1$
3. if $A_1, A_2, ...$ is a disjoint sequence of $\mathscr{F}-$sets, and if $\cup_{k=1}^\infty A_k \in \mathscr{F}$, then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \tag{3.1}$$

Property (3.1) is known as **countable additivity**. Remember that $\mathscr{F}$ is just an algebra, not necessarily a σ-algebra, so it is necessary to make the assertion that $\cup_k A_k \in \mathscr{F}$. However, we have another more constrained version of countable additivity known as **finite additivity** – that is,

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k). \tag{3.2}$$

If $\mathscr{F}$ is a σ-algebra in Ω and P is a probability measure on $\mathscr{F}$, then the triplet $(\Omega, \mathscr{F}, P)$ is a **probability measure space** or a **probability space**. An $\mathscr{F}$-set A such that $P(A) = 1$ is the **support** of the space.

From this definition, we can establish several key conclusions from classic discrete probability theory, such as **monotonicity**, the **inclusion-exclusion principle**, and **Boole's inequality** (the **union bound**, generalized here as **finite subadditivity**):

$$P\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} P(A_k) \tag{3.3}$$

3.4 The Monotone Class Theorem

Note

From here onwards, our study switches from Billingsley's text to Bass. We no longer refer to spaces as Ω and instead use X as a generalized space and discontinue much of the probabilistic interpretation that we have constructed thusfar to construct our definitions. We previously used $\mathscr{F}$ to denote an algebra; we now use the much more appropriate $\mathcal{A}$.

We can extend the properties of an algebra $\mathcal{A}$ to the σ-algebra $\mathcal{A}$ generated by $\mathcal{A}_0$

Monotone Class

A **montone class** is a collection of subsets $\mathcal{M}$ of a set X such that

1. if $A_i \uparrow A$ and each $A_i \in \mathcal{M}$ then $A \in \mathcal{M}$.
2. if $A_i \downarrow A$ and each $A_i \in \mathcal{M}$ then $A \in \mathcal{M}$.

The intersection of monotone classes is a monotone class. We arrive at the **monotone class theorem**, a rather technical and nontrivial result:

Monotone Class Theorem

Suppose $\mathcal{A}_0$ is a σ-algebra, $\mathcal{A}$ is the smallest σ-algebra containing $\mathcal{A}_0$, and $\mathcal{M}$ is the smallest monotone class containing $\mathcal{A}_0$. Then $\mathcal{M} = \mathcal{A}$.

4 A Brief Definition of Measures

Measure is a generalization of the concept of 1-dimensional length, 2-dimensional area, 3-dimensional volume, etc.

Measure

Let X be a set, and $\mathcal{A}$ a σ-algebra consisting of subsets of X. A **measure** on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \mapsto [0, \infty]$ such that

1. $\mu(\emptyset) = 0$
2. if $A_i \in \mathcal{A}$, $i = 1, 2, ...$ are pairwise disjoint, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{4.1}$$

   This is just a statement of countable additivity, except now applied to a more general notion of measure as opposed to probability measures in the previous section.

In fact, we also extend the notion of a probability space to the triple $(X, \mathcal{A}, \mu)$ which we now call a **measure space**.

The following are then true about measures:

1. If $A, B \in \mathcal{A}$, and $A \subset B$, then $\mu(A) \leq \mu(B)$ (monotonicity)
2. If $A_i \subset \mathcal{A}$, then $\mu(\cup_{i=1}^{\infty} A_i \leq \sum_{i=1}^{\infty} \mu(A_i)$ (countable additivity).
3. Suppose $A_i \in \mathcal{A}$ and $A_i \uparrow A$. Then $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.
4. Suppose $A_i \in \mathcal{A}$ and $A_i \downarrow A$. If $\mu(A_i) < \infty$, then we have $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.

A measure μ is a **finite measure** if $\mu(X) < \infty$. A measure μ is $\sigma-$**finite** if there exist sets $E_i \in \mathcal{A}$ such that $\mu(E_i) < \infty$ for each i and $X = \cup_{i=1}^{\infty} E_i$. If μ is a finite measure then $(X, \mathcal{A}, \mu)$ is called a **finite measure space**, and if μ is a σ-finite measure, then $(X, \mathcal{A}, \mu)$ is a σ**-finite measure space**.

Let $(X, \mathcal{A}, \mu)$ be a measure space. A subset $A \subset X$ is a **null set** if there is a $B \in \mathcal{A}$ with $A \subset B$ and $\mu(B) = 0$. A need not be in $\mathcal{A}$. If $\mathcal{A}$ contains all the null sets, then $(X, \mathcal{A}, \mu)$ is a **complete measure space**. The **completion** of $\mathcal{A}$ is the smallest σ-algebra $\bar{\mathcal{A}}$ containing $\mathcal{A}$ such that $(X, \bar{\mathcal{A}}, \bar{\mu})$ is complete, where $\bar{\mu}$ is a measure on $\bar{\mathcal{A}}$ that is an extension of μ, that is, $\bar{\mu}(B) = \mu(B)$ if $B \in \mathcal{A}$.

5 Construction of Measures

Here we theorize how we may *construct* measures. This is a highly complex procedure. We want the measure m of an open interval to be that interval's length. Every open subset of the reals is a countable union of disjoint open intervals. Therefore, if $G = \cup_{i=1}^{\infty}(a_i, b_i)$, where (a_i, b_i) are disjoint, then $m(G) = \sum_{i=1}^{\infty}(b_i - a_i)$. Then for subsets $E \subset \mathbb{R}$, we let

$$m(E) = \inf\{m(G) : G \text{ open }, E \subset G\} \tag{5.1}$$

In other words, the measure of E is the smallest measure of a union of disjoint open intervals containing E. However, m is not a measure on the σ-algebra of all subsets of the reals; therefore, we must consider a strictly smaller σ-algebra. This is the basis of the **Lebesgue theory of measure**; although, it is easier to work with half-open intervals $(a, b]$.

5.1 Outer Measure

Outer Measure

Let X be a set. An **outer measure** is a function μ^* defined on the collection of all subsets of X satisfying

1. $\mu^*(\emptyset) = 0$
2. If $A \subset B$, then $\mu^*(A) \leq \mu^*(B)$
3. $\mu^*(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$ when A_i are subsets of X.

The null set of an outer measure is N where $\mu^*(N) = 0$. There is a well-defined method for constructing outer measures. Suppose $\mathcal{C}$ is a collection of subsets of X such that $\emptyset \in \mathcal{C}$ and there exist $D_1, D_2, ...$ in $\mathcal{C}$ such that $X = \cup_{i=1}^{\infty} D_i$. Suppose $\ell : \mathcal{C} \mapsto [0, \infty]$ with $\ell(\emptyset) = 0$. Define the outer measure μ^* as:

$$\mu^*(E) = \inf \left\{ \sum_{i=1}^{\infty} \ell(A_i) : A_i \forall i, E \subset \cup_{i=1}^{\infty} A_i \right\}. \tag{5.2}$$

5.1.1 The Lebesgue Measure

Here we briefly introduce the Lebesgue measure, which will be given more thorough treatment in a later section. Let $X = \mathbb{R}$ and let $\mathcal{C}$ be the collection of intervals of the form $(a, b]$; that is, intervals open on the left and closed on the right. Let $\ell(I) = b - a$ if $I = (a, b]$. $\ell(I)$ is the length of I. Let μ^* be defined as in (5.2). Then μ^* is an outer measure; however, it is not an outer measure on *all* subsets of $\mathbb{R}$. If, however, we restrict μ^* to a σ-algebra $\mathcal{L}$, smaller than the collection of all subsets of $\mathbb{R}$, then μ^* will be a measure on $\mathcal{L}$. This measure is the **Lebesgue measure**, and $\mathcal{L}$ is the **Lebesgue-σ-algebra**.

Let $X = \mathbb{R}$ and let $\mathcal{C}$ be the same as in section 5.1.1. Let $\alpha : \mathbb{R} \mapsto \mathbb{R}$ be an increasing, right-continuous function (i.e. $\lim_{y \to x+} \alpha(y) = \alpha(x)$). Then define μ^* as an outer measure by (5.2). If we further restrict μ^* to a smaller σ-algebra yields the **Lebesgue-Stieltjes measure** corresponding to α. The Lebesgue measure is a case of the Lebesgue-Stieltjes measure where $\alpha(x) = x$.

5.1.2 The Carathéorody Theorem

If μ^* is an outer measure, we call $A \subset X$ **μ^*-measurable** if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c). \tag{5.3}$$

for all $E \subset X$.

Carathéodory Theorem

If μ^* is an outer measure on X, then the collection $\mathcal{A}$ of μ^*-measurable sets is a σ-algebra. If μ is the restriction of μ^* to $\mathcal{A}$, then μ is a measure. Moreover, $\mathcal{A}$ contains all the null sets.

5.2 Lebesgue-Stieltjes Measure

Let $X = \mathbb{R}$ and let $\mathcal{C}$ be a collection of half-open intervals, with $\alpha(x)$ an increasing right-continuous function. Let $\ell((a, b]) = \alpha(b) - \alpha(a)$. Then the Lebesgue-Stieltjes measure, m^*, is

$$m^*(E) = \inf \left\{ \sum_{i=1}^{\infty} \ell(A_i) : A_i \in \mathcal{C} \forall i, E \subset \cup_{i=1}^{\infty} A_i \right\} \tag{5.4}$$

Then m^* is an outer measure, and by the Carathéodory theorem, m^* is a measure on the collection of m^*-measurable sets. If K and L are two adjacent intervals, i.e. $K = (a, b]$ and $L = (b, c]$ then $K \cup L = (a, c]$, and $\ell(K) + \ell(L) = \alpha(c) - \alpha(a) = \ell(K \cup L)$.

Let's take a step back and make sure that the measure of $(a, b]$ is appropriate. Let J_k be a finite collection of half-open finite intervals covering a finite closed interval $[C, D]$. Then

$$\sum_{k=1}^{n} [\alpha(b_k) - \alpha(a_k)] \geq \alpha(D) - \alpha(C). \tag{5.5}$$

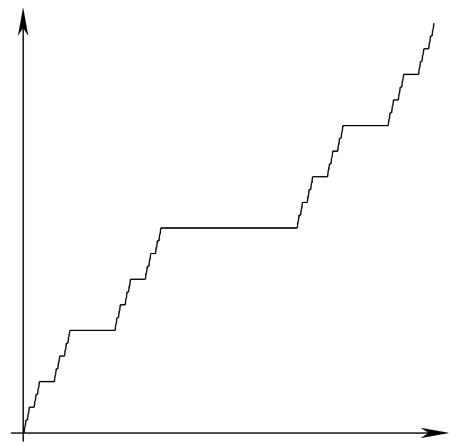
Furthermore, if a and b are finite and $I = (a, b]$, then $m^*(I) = \ell(I)$.

In order to continue our construction of the Lebesgue-Stieltjes measure, we make the following proposition: every set in the Borel σ-algebra on $\mathbb{R}$ is m^*-measurable. Dropping the asterix from m^*, we retrieve m, the **Lebesgue-Stieltjes measure**, known as the **Lebesgue measure** when $\alpha(x) = x$. In the latter case, the collection of m^*-measurable sets is $\mathcal{L}$, the **Lebesgue σ-algebra**; a Lebesgue measurable set must be $\mathcal{L}$.

5.3 Examples and the Cantor-Lebesgue Function

Recall that the **Cantor set** is constructed as follows. Lett $F_0 = [0, 1]$. Then $F_1 = F_0 - (1/3, 2/3)$. Then $F_2 = F_1 - (1/9, 2/9) \cup (7/9, 8/9)$, and continue to remove middle thirds. The Lebesgue measure of the Cantor set is 0. Let f_0 be 1/2 on the interval $(1/3, 2/3)$; let it be 1/4 on $(1/9, 2/9)$ and 3/4 on $(7/9, 8/9)$, etc. Define the following function, also known as the "devil's staircase":

$$f(x) = \inf\{f_0(y) : y \geq x, y \notin C, x < 1\},\ f(1) = 1$$



This function is constant on C^c; but since C has measure 0, then any interval in $[0, 1]$ must be in C^c since all open intervals have positive measure, and thus f must be constant everywhere and therefore continuous! This function is the **Cantor-Lebesgue function**, also called the **Cantor function**.

Instead, we could have chosen to remove the middle 1/4 – then 1/16, 1/64 and so on. Then the total removed would have been 1/4 + 2/16 + 4/64... which is 1/2! However, just like the normal Cantor set, this set has no intervals, is closed, has every point as a limit point, and is uncountable; however, it has nontrivial measure. This is the **general Cantor set**.

We sometimes call the countable intersection of open sets G_δ, German for *geoffnet Durschnitt* (open intersection), and the countable union of closed sets as F_σ (from the French *fermé* (closed) and the German *Summe* (union)). Then in understanding Lebesgue measure, we really just have to look at G_δ and F_σ sets.

5.4 The Carathéodory Extension Theorem

Let $\mathcal{A}_0$ be an algebra (but not necessarily a $\sigma-$algebra). Then define a measure on $\mathcal{A}$ as $\ell : \mathcal{A}_0 \mapsto [0, \infty]$. Let

$$\mu^*(E) = \inf \left\{ \sum_{i=1}^{\infty} \ell(A_i) : A_i \in \mathcal{A}_0, E \subset \cup_{i=1}^{\infty} A_i \right\}.$$

Then:

1. μ^* is an outer measure
2. $\mu^*(A) = \ell(A)$ if $A \in \mathcal{A}_0$
3. Every set in $\mathcal{A}_0$ and every μ^*-null set is μ^*-measurable.
4. If ℓ is σ-finite, then there is a unique extension to $\sigma(\mathcal{A}_0)$.

6 Measurable Functions

6.1 Measurability

Suppose we have $(X, \mathcal{A})$, a measurable space.

Measurable Function

$f : X \mapsto \mathbb{R}$ is **measurable** if $\{x : f(x) > a\} \in \mathcal{A}$ for all $a \in \mathbb{R}$. Complex functions require both the real and imaginary parts to be measurable. We can equivalently take $\{x : f(x) < a\}$ and $\{x : f(x) \leq a\}$ and $\{x : f(x) \geq a\}$. If X is a metric space, $\mathcal{A}$ contains all the open sets, and f is continuous, then f is measurable. Lastly, if f_i are a set of measurable functions, then $\sup_i f_i$, $\inf_i f_i$, $\limsup_i f_i$, and $\liminf_i f_i$ are all measurable if they exist.

If for two different functions f and g, $\{x : f(x) \neq g(x)\}$ has measure 0, then we say $f = g$ **almost everywhere** (they agree at all but a countable number of points). If a function is measurable on X with respect to the Borel σ-algebra $\mathcal{B}$, we say f is **Borel measurable**; from the above definition, all continuous functions are Borel measurable.

6.2 Approximation of Functions

Characteristic

Let $(X, \mathcal{A})$ be a measure space. The **characteristic function** of $E \in \mathcal{A}$ is

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

A **simple function s** is a function of the form

$$\mathbf{s}(x) = \sum_{i=1}^{n} a_i \chi_{E_i}(x).$$

If f is a non-negative measurable function, then there is a sequence of non-negative measurable simple functions s_n increasing to f.

6.3 Lusin's Theorem

The following theorem is pretty but not particularly useful.

Lusin's Theorem

Suppose $f : [0,1] \mapsto \mathbb{R}$ is Lebesgue measurable, m is Lebesgue measure, and $\varepsilon > 0$. Then there is a closed set $F \subset [0,1]$ such that $m([0,1] - F) < \varepsilon$ and the restriction of f to F is a continuous function on F.

7 The Lebesgue Integral

Let $(X, \mathcal{A}, \mu)$ be a measure space. Then the **Lebesgue integral** of the simple function

$$\mathbf{s}(x) = \sum_{i=1}^{n} a_i \chi_{E_i}(x)$$

is defined as

$$\int s d\mu = \sum_{i=1}^{n} a_i \mu(E_i). \tag{7.1}$$

If f is a measurable function, then

$$\int f d\mu = \sup \left\{ \int \mathbf{s} d\mu : 0 \leq \mathbf{s} \leq f \right\} \tag{7.2}$$

If f is measurable and $\int |f| d\mu$ is finite, then f is **integrable**.

8 Limit Theorems

8.1 Monotone Convergence Theorem

Monotone Convergence Theorem

Suppose f_n is a sequence of non-negative measurable functions with $f_1(x) \leq f_2(x) \leq \ldots$ with

$$\lim_{n \to \infty} f_n(x) = f(x) \tag{8.1}$$

Then $\lim \int f_n d\mu = \int f d\mu$.

8.2 Linearity of the Lebesgue Integral

Just like the Riemann integral, the Lebesgue integral is additive, in that

$$\int (f+g) d\mu = \int f d\mu + \int g d\mu. \tag{8.2}$$

1. If f is real-valued, measurable, and bounded and $\mu(X) < \infty$, then $a\mu(X) \leq \int f d\mu \leq b\mu(X)$.
2. If f and g are measurable, real-valued, and integrable and $f(x) \leq g(x)$ for all x, then $\int f d\mu \leq \int g d\mu$.
3. If f is complex-valued and integrable and c is a complex number, then $\int cf d\mu = c \int f d\mu$.
4. If $\mu(A) = 0$ and f is measurable, then $\int f \chi_A d\mu = 0$.

Assume f_n are non-negative measurable functions. Then

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n. \tag{8.3}$$

If f is integrable,

$$\left| \int f \right| \leq \int |f|.$$

8.3 Fatou's Lemma

Fatou's Lemma

Suppose the f_n are non-negative and measurable. Then

$$\int \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int f_n. \tag{8.4}$$

8.4 Dominated Convergence Theorem

Dominated Convergence Theorem

Suppose that f_n are measurable real-valued functions and $f_n(x) \to f(x)$ for each x. Suppose there exists a non-negative integrable function g such that $|f_n(x)| \leq g(x)$ for all x. Then

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu. \tag{8.5}$$

In both the MCT and DCT, the result still holds if $f_n \to f$ only almost everywhere (this means the set of points $\{x : f_n(x) \not\to f(x)\}$ has measure 0).

9 Properties of Lebesgue Integrals

9.1 Criteria for a function to be zero a.e.

Suppose f is measurable and non-negative and $\int f d\mu = 0$. Then $f = 0$ a.e.

Suppose f is real=valued and integrable and for every measurable set A we have $\int_A f d\mu = 0$. Then $f = 0$ a.e.

Let m be Lebesgue measure and $a \in \mathbb{R}$. Suppose $f : \mathbb{R} \mapsto R$ is integrable and $\int_a^x f(y)dy = 0$ for all x. Then $f = 0$ a.e.

9.2 An approximation result

Suppose f is a Lebesgue measurable real-valued integrable function on $\mathbb{R}$. Let $\varepsilon > 0$. Then there exists a continuous function g with compact support such that

$$\int |f - g| < \varepsilon. \tag{9.1}$$

10 The Riemann Integral

Here we compare the Lebesgue and Riemann integrals. First, recall the Darboux construction of the definition of the Riemann integral. As we have reserved the integral symbol $\int$ for the Lebesgue integral, we use $R(f)$ to denote the Riemann integral of f. Let P be a partition $\{x_0, x_1, ..., x_n\}$ of the region of integration $[a, b]$ with $x_0 = a$ and $x_n = b$. Then the upper Darboux sum is

$$U(P, f) = \sum_{i=1}^{n} \left(\sup_{x_{i-1} \leq x \leq x_i} f(x) \right) (x_i - x_{i=1}).$$

The lower Darboux sum is similarly

$$L(P, f) = \sum_{i=1}^{n} \left(\inf_{x_{i-1} \leq x \leq x_i} f(x) \right) (x_i - x_{i=1}).$$

The upper and lower Darboux integrals are then

$$\overline{R}(f) = \inf\{U(P, f)\}, \quad \underline{R}(f) = \sup\{L(P, f)\}$$

over all partitions P. The Riemann integral $R(f)$ exists if $\overline{R}(f) = \underline{R}(f)$, and the value they attain is the value of $R(f)$.

10.1 Comparison with the Lebesgue integral

A bounded real-valued function f on $[a, b]$ is Riemann integrable if and only if the set of points at which f is discontinuous has Lebesgue measure 0, and in that case, f is Lebesgue measurable and the Riemann integral of f is equal in value to the Lebesgue integral of f.

11 Types of convergence

Almost everywhere convergence

If μ is a measure, we say a sequence of measurable functions f_n converges **almost everywhere** to f and write $f_n \to f$ a.e. if the set $\{x : f_n(x) \not\to f(x)\}$ has measure 0.

f_n **converges in measure** to f if for all $\varepsilon > 0$,

$$\mu(\{x : |f_n(x) - f(x)| > \varepsilon\}) \to 0 \tag{11.1}$$

as $n \to \infty$. For $1 \leq p < \infty$, we say f_n **converges in L^p to** f if

$$\int |f_n - f|^p d\mu \to 0 \tag{11.2}$$

as $n \to \infty$.

11.1 Chebyshev's Inequality

Here we show how convergence in L^p corresponds to convergence in measure.

Chebyshev's Inequality

$$\mu(\{x : |f(x)| \geq a\}) \leq \frac{\int |f|^p d\mu}{a^p}. \tag{11.3}$$

As a result, if f_n converges to f in L^p, then it converges in measure.

11.2 Egorov's Theorem

Suppose μ is a finite measure, $\varepsilon > 0$ and $f_n \to f$ a.e. Then there exists a measurable set A such that $\mu(A) < \varepsilon$ and $f_n \to f$ uniformly on A^c. This kind of convergence is known as **almost uniform convergence**, but it is not particularly useful.

12 Signed Measures

Signed measures are measures which are allowed to take both negative and positive values.

12.1 Positive and Negative Sets

Signed Measure

Let $\mathcal{A}$ be a σ-algebra. A **signed measure** is a function $\mu : A \mapsto (-\infty\infty)$ such that $\mu(\emptyset) = 0$ and if $A_1, A_2, ...$ are pairwise disjoint and all the A_i are in $\mathcal{A}$ then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$, where the series converges absolutely if $\mu(\cup_{i=1}^{\infty} A_i)$ is finite.

A set $A \in \mathcal{A}$ is a **positive set** for μ if $\mu(B) \geq 0$ for $B \subset A$ and $B \in \mathcal{A}$. $A \in \mathcal{A}$ is a **negative set** if $\mu(B) \leq 0$ for $B \subset A$ and $B \in \mathcal{A}$. If $\mu(B) = 0$ for $B \subset A$ and $B \in \mathcal{A}$, then A is a **null set**.

12.2 Hahn Decomposition

Recall that $A \triangle B = (A-B) \cup (B-A)$. The **Hahn decomposition theorem** asserts that if X has a signed measure μ, then X can be decomposed into a positive set where μ acts as a positive measure and a negative set where $-\mu$ acts as a positive measure.

1. Let μ be a signed measure taking values in $(-\infty, \infty)$. There exist disjoint measurable sets E and F in $\mathcal{A}$ whose union is X and such that E is a negative set and F is a positive set.

2. If E' and F' are another such pair, then $E \triangle E' = F \triangle F'$ is a null set with respect to μ; thus the decomposition is "unique" (modulo null sets).

3. If μ is not a positive measure, then $\mu(E) < 0$. If $-\mu$ is not a positive measure, then $\mu(F) > 0$.

Two measures μ, ν are **mutually singular** if there exist disjoint sets E, F in $\mathcal{A}$ whose union in X with $\mu(E) = \nu(F) = 0$; this is written as $\mu \perp \nu$.

12.3 Jordan Decomposition

If μ is a signed measure on a measurable space $(X, \mathcal{A})$, there exist positive measures μ^+ and μ^- such that $\mu = \mu^+ - \mu^-$ and μ^+ and μ^- are mutually singular. This decomposition is unique.

Proof: Let E and F be negative and positive sets, respectively for a signed measure μ such that $X = E \cup F$ and $E \cap F = \emptyset$. Let $\mu^+(A) = \mu(A \cap F)$, $\mu^-(A) = -\mu(A \cap E)$. This yields the desired decomposition.

Imagine that $\mu = \nu^+ - \nu^-$ is a second, distinct decomposition from the one described, with ν^+, ν^- mutually singular. Let E' be such that $\nu^+(E') = 0$ and $\nu^-(E'^c) = 0$. Let $F' = E'^c$. Then $X = E' \cup F'$ and $E' \cap F' = \emptyset$. If $A \subset F'$then $\nu^-(A) \leq \nu^-(F) = 0$, so $\nu^+(A) \geq 0$ and F' is a positive set; similarly E' is a negative set. Then E' and F' gives a Hahn decomposition of X. As the Hahn decomposition is unique modulo null sets, $F \triangle F'$ is a null set with respect to μ, as is $E \triangle E'$. Now since $\nu^+(E') = \nu^-(F') = 0$, we must have $\nu^+ = \mu^+$ and $\nu^- = \mu^-$ acting on sets $A \in \mathcal{A}$. $\square$

13 The Radon-Nikodym Theorem

So far, given countable additivity, we can say definitively that

$$\nu(A) = \int_A f d\mu \tag{13.1}$$

is a measure. Here we examine a theorem that informs us of the converse; that, given μ and ν, when does there exist an f such that (13.1) holds?

13.1 Absolute Continuity

A measure ν is **absolutely continuous** with respect to μ if $\nu(A) = 0$ whenever $\mu(A) = 0$. We write $\nu \ll \mu$. More formally, $\nu \ll \mu$ if and only if for all ε there exists δ such that $\mu(A) < \delta \implies \nu(A) < \varepsilon$.

13.2 Main Theorem

Radon-Nikodym Theorem

Suppose μ is a σ-finite positive measure on a measurable space $(X, \mathcal{A})$ and ν is a finite positive measure on $(X, \mathcal{A})$ such that ν is absolutely continuous with respect to μ. Then there exists a μ-integrable non-negative function f which is measurable with respect to $\mathcal{A}$ such that

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

for all $A \in \mathcal{A}$. If g is another such function, then $f = g$ a.e. w.r.t μ.

f is the **Radon-Nikodym derivative** of ν w.r.t μ (sometimes the **density** of ν w.r.t μ), written $f = \frac{d\nu}{d\mu}$. A common formulation is

$$d\nu = f d\mu. \tag{13.2}$$

13.3 Lebesgue Decomposition

The **Lebesgue decomposition theorem** gives another decomposition of measures. Suppose μ is a σ-finite positive measure and ν is a finite positive measure. Then there exist positive measures λ, ρ such that $\nu = \lambda + \rho$, $\rho \ll \mu$, $\lambda \perp \mu$.

14 Differentiation

Our approach to differentiation uses what are known as maximal functions in combination with the Radon-Nikodym theorem and Lebesgue decomposition.

The definition of derivative is from elementary calculus. f is **differentiable** at x if

$$\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \tag{14.1}$$

exists; the limit is called the **derivative of** f **at** x and is denoted $f'(x)$. f is differentiable on $[a, b]$ if $f'(x)$ exists for all $x \in (a, b)$ and the positive and negative limits of (14.1) exist. The ultimate goal of this section is to construct the Fundamental Theorem of Calculus.

14.1 Hardy-Littlewood Maximality

Consider real-valued functions on $\mathbb{R}^n$. Let $B(x, r)$ be the open ball with center x and radius r. The following proposition is what we call a **covering lemma**.

Suppose $E \subset \mathbb{R}^n$ is covered by a collection of balls $\{B_\alpha\}$ and there exists a positive real number R such that the diameter of each B_α is bounded by R. Then there exists a disjoint sequence $B_1, B_2, ...$ of elements of $\{B_\alpha\}$ such that

$$m(E) \leq 3^n \sum_k m(B_k). \tag{14.2}$$

f is **locally integrable** if $\int_K |f(x)|dx$ is finite for K compact. If f is locally integrable, define

$$Mf(x) = \sup_{r>0} \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y)|dy. \tag{14.3}$$

The function Mf is the **maximal function** of f, and M is known as the **Hardy-Littlewood maximal operator**.

Then we get the main inequality of the section; a **weak 1-1 inequality** (also attributed to Hardy and Littlewood). This inequality does not map integrable functions into integrable functions, but (in a way) comes close to doing so.

$$m(\{x : Mf(x) > \beta\}) \leq \frac{3^n}{\beta} \int |f(x)|dx \tag{14.4}$$

Note that Mf for $f = \chi_B$ is not integrable, showing the weakness of the inequality.

There are two further results, of increasing strength, that help us determine the first half of the Fundamental Theorem of Calculus. The weaker: Let

$$f_r(x) = \frac{1}{m(B(x,r))} \int_{B(x,r)} f(y)dy \tag{14.5}$$

If f is locally integrable, then $f_r \to f$ a.e. as $r \to 0$. The stronger: for almost every x,

$$\frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|dy \to 0 \tag{14.6}$$

as $r \to 0$.

14.2 Antiderivatives

Here we determine that the derivative of the antiderivative of an integrable function is the function itself. Let $B(x,h)$ in $\mathbb{R}$ be the interval $(x-h, x+h)$ (the h-ball about x). m is Lebesgue measure as is standard. The **indefinite integral** or **antiderivative** of an integrable function f is

$$F(x) = \int_a^x f(t)dt. \tag{14.7}$$

F is then differentiable a.e. and $F'(x) = f(x)$ a.e.

14.3 Increasing Functions

Here we assert that increasing functions are differentiable almost everywhere, starting with right continuity.

Suppose $H : \mathbb{R} \mapsto \mathbb{R}$ is increasing, right continuous, and constant for $x \geq 1$ and $x \leq 0$. Let λ be the Lebesgue-Stieltjes measure defined using the function H and suppose λ and m are mutually singular. Then

$$\lim_{r \to 0} \frac{\lambda(B(x,r))}{m(B(x,r))} = 0$$

for almost every x with respect to m.

Now let $F : \mathbb{R} \mapsto \mathbb{R}$ be an increasing and right continuous function. Then F' exists a.e. Moreover, F' is locally integrable and for every $a < b$, $\int_a^b F'(x)dx \leq F(b) - F(a)$. In fact, we can further generalize this and remove our requirement for right continuity, sacrificing local integrability; in other words, an increasing function has a derivative which exists a.e. such that

$$\int_a^b F'(x)dx \leq F(b) - F(a) \tag{14.8}$$

when $a < b$.

14.4 Bounded Variation

A real-valued function f is of **bounded variation** on $[a, b]$ if

$$V_f[a, b] = \sup \left\{ \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})| \right\} \tag{14.9}$$

f is **Lipschitz continuous** if there exists $c_1 > 0$ such that

$$|f(y) - f(x)| \leq c_1 |y - x| \tag{14.10}$$

for all x, y. If f is Lipschitz continuous then f must also be of bounded variation.

The most important result here is the following: if f is of bounded variation on $[a, b]$, then f can be written as $f = f_1 - f_2$, where f_1 and f_2 are increasing functions on $[a, b]$. This, along with (14.8), shows that functions of bounded variation are differentiable a.e.

14.5 Absolutely Continuous Functions

Absolutely Continuous Function

A real-valued function is **absolutely continuous** on $[a, b]$ if given ε there exists δ such that $\sum_{i=1}^{k} |f(b_i) - f(a_i)| < \varepsilon$ whenever $\{(a_i, b_i)\}$ is a finite collection of disjoint intervals with $\sum_{i=1}^{k} |b_i - a_i| < \delta$. If f is absolutely continuous, then it is of bounded variation.

Furthermore, if we decompose f into $f_1 - f_2$, as in the previous subsection, then f_1 and f_2 are also absolutely continuous.

If F is absolutely continuous, then F' exists a.e., is integrable, and

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

Proof: Suppose F is increasing and absolutely continuous. Let ν be the Lebesgue-Stieltjes measure. Since F is continuous, $F(d) - F(c) = \nu((c,d))$. Letting $k \to \infty$, we see given ε there exists a δ such that $\sum_{i=1}^k |F(b_i) - F(a_i)| < \varepsilon$ whenever $\{(a_i, b_i)\}$ is a finite collection of disjoint intervals with $\sum_{i=1}^k |b_i - a_i| < \delta$. Since any open set G is the union of disjoint intervals $\{(a_i, b_i)\}$, we can rewrite this as

$$\nu(G) = \sum_{i=1}^{\infty} \nu((a_i, b_i)) = \sum_{i=1}^{\infty} (F(b_i) - F(a_i)) \leq \varepsilon$$

whenever G is open with $m(G) < \delta$. If $m(A) < \delta$ and A is Borel measurable, then there exists an open G containing A such that $m(G) < \delta$ and $\nu(A) \leq \varepsilon$ so $\nu \ll m$. Then by Radon-Nikodym there exists a non-negative integrable f such that

$$\nu(A) = \int_A f dm$$

for all Borel measurable sets A. In particular,

$$F(x) - F(a) - \nu((a,x)) = \int_a^x f(y)dy$$

By (14.7) $F' = f$ a.e. Setting $x = b$ we have

$$F(b) - F(a) = \int_a^b F'(y)dy.$$

□

15 L^p Spaces (Lebesgue Spaces)

15.1 Norms, Hölder's Inequality, Minkowski's Inequality

Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space. For $1 \leq p \leq \infty$, the L^p **norm** of f is

$$\|f\|_p = \left(\int |f(x)|^p d\mu \right)^{1/p}. \tag{15.1}$$

If $p = \infty$ then the L^∞ **norm** is

$$\|f\|_\infty = \inf\{M \geq 0 : \mu(\{x : |f(x)| \geq M\}) = 0\}. \tag{15.2}$$

In other words, the L^∞ norm of a function is the smallest value which upper bounds f a.e.

The L^p **space** is the set $\{f : \|f\|_p < \infty\}$. If $1 \leq p \leq \infty$, then the **conjugate exponent** of p is the number q such that $1/p + 1/q = 1$.

Hölder's Inequality

If $1 < p$, $q < \infty$, $p^{-1} + q^{-1} = 1$, and f and g are measurable, then

$$\int |fg| d\mu \leq \|f\|_p \|g\|_q. \tag{15.3}$$

This also holds if $p = \infty$ and $q = 1$ or $p = 1$ and $q = \infty$. If $p = q = 2$, this is the **Cauchy-Schwarz inequality**.

This conclusion allows us to arrive at **Minkowski's Inequality**, a generalized version of the triangle inequality.

Minkowski's Inequality

If $1 \leq p \leq \infty$ and f, g are measurable, then

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Here, we may be tempted to call L^p a **normed linear (vector) space** (as a brief review, a normed vector space is a vector space with a norm defined on it such that $\|x\| \geq 0$ with $\|x\| = 0 \iff x = 0$, $\|cx\| = |c|\|x\|$, and $\|x + y\| \leq \|x\| + \|y\|$, for all $x, y \in X$). The reason we *cannot* do this is because the $\|f\|_p$ being 0 need not imply that f is 0, only that f is 0 almost everywhere. To circumvent this, we define an equivalence relation – we call two functions equivalent if they differ on a set of measure 0, and L^p is the set of appropriate equivalence classes.

15.2 Completeness

If $1 \leq p \leq \infty$, then L^p is complete. Furthermore, the set tof continuous functions with compact support is dense in $L^p(\mathbb{R})$ for $1 \leq p \leq \infty$. Finally, the set of continuous functions on $[a, b]$ are dense in $L^2([a, b])$ with respect to the $L^2([a, b])$ norm.

15.3 Convolutions

Assume all functions here are on $\mathbb{R}^n$ with the Lebesgue measure on $\mathbb{R}^n$. The **convolution** of two measurable functions f and g is defined by

$$f * g(x) = \int f(x-y)g(y)dy$$

provided the integral exists.
Suppose f and g are Borel measurable; if $a \in \mathbb{R}$ then $A = f^{-1}((a,\infty))$ is a Borel measurable subset of $\mathbb{R}$. Then $K : \mathbb{R}^2 \to \mathbb{R}$, $K(x,y) = x - y$ is Borel measurable. Then with $X = \mathbb{R}^2$ and the Borel $\sigma-$algebra on $\mathbb{R}^2$, $K^{-1}(A) \in \mathcal{A}$ so

$$(f \circ K)^{-1}((a,\infty)) = K^{-1}(f^{-1}((a,\infty))) = K^{-1}(A)$$

is a Borel subset of $\mathbb{R}^2$; we conclude that $f \circ K$ is Borel measurable, as is g. As the integrand is jointly measurable, $f * g$ is a Borel measurable function of x. By change of variables, $f * g = g * f$.

1. If $f, g \in L^1$, then $f * g \in L^1$ and

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1.$$

2. If $1 \leq p \leq \infty$, $f \in L^1$ and $g \in L^p$ then

$$\|f * g\|_p \leq \|f\|_1 \|g\|_p.$$

We can use convolutions to approximate functions in L^p by smooth functions, a process known as **mollification**. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be infinitely differentiable with compact support, non-negative, and integrating to 0. Let

$$\varphi_\varepsilon(x) = \varepsilon^{-n}\varphi(x/\varepsilon).$$

$f * \varphi_\varepsilon$ is infinitely differentiable. For non-negative integers $\{\alpha_k\}_{k=1}^n$:

$$\frac{\partial^{\sum_k \alpha_k}(f * \varphi_\varepsilon)}{\prod_k \partial x_k^{\alpha_k}} = f * \frac{\partial^{\sum_k \alpha_k}\varphi_\varepsilon}{\prod_k \partial x_k^{\alpha_k}}.$$

$f * \varphi_\varepsilon \to f$ a.e. as $\varepsilon \to 0$. If f is continuous, then $f * \varphi_\varepsilon \to f$ uniformly on compact sets as $\varepsilon \to 0$. Finally, $f * \varphi_\varepsilon$ in L^p.

16 Topology

This is the final section of this course, which deals with general topological spaces, compactness, connectedness, separation, embeddings, and approximation. We begin this (very large) section with an extensive set of definitions.

Topology

Let X be a set. A **topology** $\mathcal{T}$ is a collection of subsets of X such that

1. $X, \emptyset \in \mathcal{T}$
2. If $G_\alpha \in \mathcal{T}$ for each α in a non-empty indexing set I, then $\cup_{\alpha \in I} G_\alpha \in \mathcal{T}$ (the countable union of the elements of $\mathcal{T}$ is in $\mathcal{T}$)
3. If $G_1, ...G_n \in \mathcal{T}$, then $\cap_i^n G_i \in \mathcal{T}$ (the finite intersection of the elements of $\mathcal{T}$ is in $\mathcal{T}$)

A **topological space** is a set X with a topology $\mathcal{T}$ defined on it. An element $G \in \mathcal{T}$ is called an **open set**; a set F is **closed** if F^c is open. If $\mathcal{T}$ consists of all subsets of X, it is the **discrete topology**. If $\mathcal{T} = \{\emptyset, X\}$, it is the **trivial topology**.

Say we have a metric space (X, d). Then $G \subset X$ is oipen in (X, d) if, when $x \in G$, there exists r_x such that $B(x, r_x) \subset G$ with $B(x, r_x) = \{y : d(x, y) < r_x\}$. If $\mathcal{T}$ is the collection of open sets, $(X, \mathcal{T})$ is a topological space generated by the metric d.

We now re-introduce some ideas from elementary analysis. Let $A \subset (X, \mathcal{T})$ (note this does not mean $A \in \mathcal{T}$). Then $x \in X$ is an **interior point** of A if $x \subset G \subset A$. The set of interior points of A is A^o, the **interior** of A.

A point x, not necessarily an element of A, is a **limit point** of A if every open set that contains x contains a point of A other than x. The set of limit points is sometimes denoted A'. The **closure** of A is $\bar{A} = A \cap A'$. The **boundary** of A, written ∂A, is $\bar{A} - A^o$. x is an **isolated point** if $x \in A - A'$ (a point which is not a limit point).

As an example, let X be the real line. Let the topology $\mathcal{T}$ be generated from the usual metric $d(x,y) = |x-y|$. Let $A = (0,1]$. Then $A^o = (0,1)$, $A' = [0,1]$, $\bar{A} = [0,1]$, and $\partial A = \{0,1\}$.

A set in A is a **neighborhood** of x if $x \in A^o$. If A is both a neighborhood *and* and open set, it is called an **open neighborhood**.

Now we discuss the case of multiple topologies. Let $(X, \mathcal{T})$ be a topological space and let $Y \subset X$. If we let $\mathcal{U} = \{G \cap Y : G \in \mathcal{T}\}$, then if $\mathcal{U}$ satisfies the definitions applied to subsets of Y, $(Y,\mathcal{U})$ is a **subspace** of $(X, \mathcal{T})$. $\mathcal{U}$ is the **relative topology**.

Given two topologies $\mathcal{T}$ and $\mathcal{T}'$, with $\mathcal{T} \subset \mathcal{T}'$, we call $\mathcal{T}$ a **weaker** or **coarser** topology and $\mathcal{T}'$ a **stronger** or **finer** topology.

Suppose $(X, \mathcal{T})$ is a topological space and $\sim$ an equivalence relation for X. Let $\bar{X}$ be the set of equivalence classes, and let $E : X \to \bar{X}$ map an element to its equivalence class. Then the **quotient topology** is
$\mathcal{U} = \{A \subset \bar{X} : E^{-1}(A) \in \mathcal{T}\}$.

A subcollection $\mathcal{B}$ of $\mathcal{T}$ is an **open base** if every element of $\mathcal{T}$ is a union of sets in $\mathcal{B}$. A subcollection $\mathcal{S}$ of $\mathcal{T}$ is a **subbase** if the collections of finite intersections of elements in $\mathcal{S}$ is an open base of $\mathcal{T}$.

Any collection $\mathcal{C}$ of subsets of X generates a topology $\mathcal{T}$ on X, with $\mathcal{T}$ being the smallest topology with $\mathcal{C}$ as a subbase. Suppose I is an index set and for each $\alpha \in I$, $(X_\alpha, \mathcal{T}_\alpha)$ is a topological set. Let $X = \prod_{\alpha \in I} X_\alpha$, and let π_α be the projection of X onto X_α. Then the topology $\mathcal{C}_\alpha = \cup_{\alpha \in I}\{\pi_\alpha^{-1}(A) : A \in \mathcal{T}_\alpha\}$ is the **product topology**.

A subcolletion $\mathcal{B}_x$ of open sets containing the point x is an **open base at point** x if every open set containing x contains an element of $\mathcal{B}_x$. If the closure of A has empty interior, A is **nowhere dense**. X is separable if there exists a countable subset of X that is dense in X. If X has a countable *base*, it is called **second countable**. A topological space is **first countable** if every point x has a countable open base at x. For a metric space X, X is second countable if and only if it is separable.

I is a **directed set** if it has an ordering $\leq$ that satisfies identity, transitivity, and upper bound. A **net** is a mapping from a directed set I into a topological space X. A net $\langle x_\alpha \rangle, \alpha \in I$ **converges** to a point y if, for each open set G containing y, there is an $\alpha_0 \in I$ such that $x_\alpha \in G$ whenever $\alpha \geq \alpha_0$. If there is a net with infinitely many points in $E \subset (X, \mathcal{T})$ that converges to y, then y is a limit point of E. Note that there is no concept of "almost everywhere" convergence in topology.

We conclude this set of definitions by discussing continuous functions.

Homeomorhpism

Suppose we have two topological spaces, $(X, \mathcal{T})$ and $(Y, \mathcal{U})$. Then $f : X \to Y$ is **continuous** if $f^{-1}(G) \in \mathcal{T}$ for $G \in \mathcal{U}$. f is **open** if $f(H) \in \mathcal{U}$ if $H \in \mathcal{T}$. f is called a **homeomorphism** if it is one-to-one, onto, continuous, and open. In this case, openness is equivalent to saying that f^{-1} is continuous.

16.1 Compactness

An open cover of $A \subset (X, \mathcal{T})$ is a non-empty collection of open subsets, $\{G_\alpha\}$, such that $A \subset \cup_{\alpha \in I} G_\alpha$. A **subcover** is a collection of $\{G_\alpha\}$ that also covers A. A is **compact** if every cover of A has a finite subcover. A closed subset of a compact set is compact.

A set A is **precompact** if $\bar{A}$ is compact. A is **σ-compact** if there exist $K_1, K_2, ...$ compact such that $A = \cup K_i$. A is countably compact if every countable cover of A has a finite subcover.

A is **sequentially compact** if every sequence of elements in A has a subsequence converging to a point in A. A has the **Bolzano-Weierstrass property** if every infinite subset of A has a limit point in A.

16.2 Tychonoff's Theorem

Zorn's Lemma

If Y is a partially ordered set and every linearly ordered subset of Y has an upper bound, then Y has a maximal element. Zorn's Lemma is functionally equivalent to the axiom of choice.

Let $(X, \mathcal{T})$ be a topological space, let $\mathcal{B}$ be a basis for $\mathcal{T}$, and let $\mathcal{S}$ be a subbasis. If $A \subset X$ and $\{G_\alpha\}$ is an open cover for A such that each $G_\alpha \in \mathcal{B}$, then $\{G_\alpha\}$ is a **basic open cover**; if $G_\alpha \in \mathcal{S}$, then $\{G_\alpha\}$ is a **subbasic open cover**.

Suppose A is a subset of X and every basic open cover of A has a finite subcover. Then A is compact. Using Zorn's lemma, we can additionally show the following: let A be a subset of X. Suppose $\mathcal{C} \subset \mathcal{E}$ are two collections of open subsets of X and suppose that no finite subcollection of $\mathcal{C}$ covers A. Then there exists a maximal subset $\mathcal{D}$ of $\mathcal{E}$ that contains $\mathcal{C}$ and that no finite subcollection of $\mathcal{D}$ covers A. We can then use the existence of this maximal subset to show that, if ever subbasic open cover of A has a finite subcover, then A is compact. The culmination of all of these statement reveals the following:

Tychonoff's Theorem

The non-empty product of compact topological spaces is compact.

16.3 Compactness and Metric Spaces

Let X be a metric space with metric d. A set A is a **bounded set** if there exists $x_0 \in X$ and $M > 0$ such that $A \subset B(x_0, M)$. If A is a compact subset of am metric space X, then A is closed and bounded. Additionally, A has the Bolzano-Weierstrass property if and only if it is sequentially compact. Finally, the following statements are equivalent:

1. A is compact;
2. A is sequentially compact;
3. A has the Bolzano-Weierstrass property.

Heine-Borel Theorem

A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Given a set A, an $\varepsilon - net$ for A is a subset $\{x_1, x_2, ...\}$ such that $\{B(x_i, \varepsilon)\}$ covers A. A is **totally bounded** if for each ε there exists a finite $\varepsilon-$net. $A \subset (X, d)$ is compact if and only if it is both complete and totally bounded.

If (X, d_X) and (Y, d_Y) are metric spaces, then $f : X \to Y$ is **uniformly continuous** if given ε there exists δ such that $d_Y(f(x), f(y)) < \varepsilon$ whenever $d_X(x, y) < \delta$. If X is a compact metric space and Y is a metric space, then if $f : X \to Y$ is continuous then it is uniformly continuous.

If X and Y are metric spaces, $\varphi : X \to Y$ is an **isometry** if $d_Y(\varphi(x), \varphi(y)) = d_X(x, y)$ for all $x, y \in X$ where d_X is the metric for X and d_Y is the one for Y. A metric space X^* is the **completion** of X if there is an isometry $\varphi : X \to X^*$ such that $\varphi(X)$ is dense in X^* and X^* is complete. All metric spaces have a completion.

16.4 Separation Properties

Here we have some different kind of spaces.

A topological space X is a T_0 **space** or **Kolmogorov space** if, whenever $x \neq y$, there exists an open set G such that either $x \in G, y \notin G$ or $y \in G, x \notin G$.

X is a T_1 **space** if, whenever $x \neq y$, there exists an open set G containing x which does not contain y and an open set H containing y which does not contain x.
X is a T_2 or **Hausdorff space** if whenever $x \neq y$, there exist disjoint open sets G and H such that $x \in G$ and $y \in H$. G and H **separate** x and y.

X is a **completely regular space** or **Tykhonov space** or $T_{3\frac{1}{2}}$ space if X is a T_1 space and whenever $F \subset X$ is closed, $x \notin F$, there is a continuous real valued function such that $f(x) = 0$ and $f(y) = 1$ for $y \in F$.

X is a **normal space** or T_4 space if X is a T_1 space and whenever E and F are disjoint and closed, there exist open sets G and H such that $E \subset G$ and $F \subset H$.

16.4.1 Separation Properties

Singletons of T_1 spaces are closed. Furthermore, the product of a non-empty class of Hausdorff spaces is Hausdorff; then compact subsets of a Hausdorff space are closed. A compact Hausdorff space is normal.

16.5 Urysohn's Lemma

Urysohn's lemma shows that normal spaces have a plentiful supply of continuous functions. Disjoint closed subsets of compact Hausdorff spaces can be separated by continuous functions. Urysohn's lemma is the following:

Urysohn's Lemma

Let E and F be disjoint closed subsets of a normal space X. There exists a continuous real-valued function taking values in $[0,1]$ such that $f = 0$ on E and $f = 1$ on F.

As a corollary, if X is a compact Hausdorff space, K is a compact subset of X, and G is an open subset of X containing K, then there exists a continuous function f that is 1 on K and such that the support of f is contained in G.

16.6 Tietze Extension Theorem

Suppose $f : C \times C \to [0,1]$ is continuous (C is the middle-thirds Cantor set). Can we extend f so that $f : [0,1]^2 \to [0,1]$ is continuous?

Tietze Extension Theorem

Let X be normal, F a closed subspace, and $f : F \to [a,b]$ a continuous function. There exists a continuous function $\overline{f} : X \to [a,b]$ which is an extension of f, that is $\overline{f}\mid_F = f$.

16.7 Urysohn Embedding Theorem

Compact Hausdorff spaces can sometimes substitute for metric spaces. To be more specific, Hausdorff spaces that are second countable can be made *into* metric spaces! This is the the remark made in the Urysohn embedding theorem (equivalently the Urysohn metrization theorem). $(X, \mathcal{T})$ is **metrizable** if there is a metric d such that a set is open with respect to d if and only if it is in $\mathcal{T}$. More precisely, if $x \in G \in \mathcal{T}$, there exists an $r-$ball such that $B(x,r) \subset G$ and $B(x,r) \in \mathcal{T}$ for each x and $r > 0$.

We can embed second countable normal spaces into $[0,1]^{\mathbb{N}}$, with the metric in question being

$$d(x,y) = \sum_{i=1}^{\infty} 2^{-i}|x_i - y_i|.$$

Formally, the Urysohn embedding theorem states that for a second countable normal space X, there exists a homeomorphism φ of X onto a subset of $[0,1]^{\mathbb{N}}$; in particular, X is metrizable.

16.8 Alexandroff One-Point Compactification

A topological space is **locally compact** if each point has a neighborhood of compact closure. Let $(X, \mathcal{T})$ be a locally compact Hausdorff space. Let ∞ denote a point not in X and let $X^* = X \cup \{\infty\}$. Define $\mathcal{T}^*$ to consist of X^*, all elements of $\mathcal{T}$, and all sets $G \subset X^*$ such that G^c is compact in $(X, \mathcal{T})$. Then $(X^*, \mathcal{T}^*)$ is a compact Hausdorff space. X^* is known as the **Alexandroff one-point compactification** of X.

16.9 Stone-Cech Compactification

Given a completely regular space X, we can find a compact Hausdorff space $\overline{X}$ such that X is dense in $\overline{X}$ and every bounded continuous function on X can be extended to a bounded continuous function on $\overline{X}$. This is the **Stone-Cech compactification**[1] of X, written as $\beta(X)$. This theorem is incredibly useful; the function $f(x) = \sin(1/x)$ on $(0, 1]$ cannot be extended to $[0, 1]$, so $\beta(X) \neq [0, 1]$; however, there is a compactification of $(0, 1]$ for which f *does* have a continuous extension.

Stone-Cech Compactification

Let X be a completely regular space. There exists a compact Hausdorff space $\beta(X)$ and a homeomorphism ϕ mapping X into a dense subset of $\beta(X)$ such that if f is a bounded continuous function from X to $\mathbb{R}$, then $f \circ \varphi^{-1}$ has a bounded continuous extension to $\beta(X)$. Density ensures that only one such extension exists for each function.

16.10 Ascoli-Arzelà Theorem

This theorem is especially powerful; it is a simple way to check whether a collection of continuous functions on a compact Hausdorff space contains a sequence with a uniformly convergent subsequence. While this is common in introductory analysis, the variant here is that we allow X to be a compact Hausdorff space, and can be proven without relying on diagonalization.

Let X be a compact Hausdorff space and $\mathcal{C}(X)$ be the set of continuous functions on X. Compactness ensures $f(X)$ is compact and therefore bounded.

$$d(f, g) = \sup_{x \in X} |f(x) - g(x)|$$

[1] The "C" in Cech is accented with a caron, which unfortunately does not render in this document.

then transforms $\mathcal{C}(X)$ into a metric space.

$\mathcal{F} \subset \mathcal{C}(X)$ is **equicontinuous** if, given ε and x there is an open set G containing x such that if $y \in G$ and $f \in \mathcal{F}$, then $|f(y) - f(x)| < \varepsilon$. The same G must work for *every* $f \in \mathcal{F}$, making equicontinuity stronger than continuity.

Ascoli-Arzelà Theorem

Let X be a compact Hausdorff space and let $\mathcal{C}(X)$ be the set of continuous functions on X. Then $\mathcal{F} \subset \mathcal{C}(X)$ is compact if and only if:

1. $\mathcal{F}$ is closed;
2. $\sup_{f \in \mathcal{F}} |f(x)| < \infty$ for each $x \in X$;
3. $\mathcal{F}$ is equicontinuous.

16.11 Stone-Weierstrass Theorems

The Stone-Weierstrass theorems allow one to approximate continuous functions. We begin by introducing the **Weierstrass approximation theorem**, which allows for approximating real-valued functions on a compact interval.

Let $[a, b]$ be a finite subinterval of $\mathbb{R}$, g a continuous function on $[a, b]$ and $\varepsilon > 0$. Then there exists $P(x)$ such that

$$\sup_{x \in [a,b]} |g(x) - P(x)| < \varepsilon.$$

This can be proven using some ideas from earlier in this course. For instance, if we let $\varphi_\beta(x)$ be a Gaussian of 0-mean and β-standard deviation. Using the DCT one may show that g has compact support and is continuous. The convolution $g * \varphi_\beta(x)$ is close to g uniformly as β vanishes. Then $\varphi_\beta(x)$ can be approximated by a polynomial via Taylor series; all that remains is a change of variables to conclude.

Let X be a topological space, with $\mathcal{C}(X)$ the set of real-valued continuous functions on X. $\mathcal{A} \subset \mathcal{C}(X)$ is an **algebra of functions** if $\mathcal{A}$ is closed under addition and multiplication (note $\mathcal{A}$ is not necessarily a ring algebra, as no claim is made about unity, commutativity, or associativity). If $\mathcal{A}$ is closed under minimum and maximum, it is a **lattice of functions**. $\mathcal{A}$ **separates** $x \neq y$ if there exists $f \in \mathcal{A}$ such that $f(x) \neq f(y)$. $\mathcal{A}$ **vanishes at no point** of X if whenever $x \in X$, there exists $g \in \mathcal{A}$ with $g(x) \neq 0$.

Stone-Weierstrass Theorems

Suppose X is a compact Hausdorff space and $\mathcal{A}$ is an algebra of real-valued continuous functions that separates points and vanishes at no point. Then $\mathcal{A}$ is dense in $\mathcal{C}(X)$.

Now suppose X is a compact Hausdorff space and $\mathcal{C}(X, \mathbb{C})$ is the set of complex-valued continuous functions on X. Let $\mathcal{A}$ be an algebra of continuous complex-valued functions that separates points and vanishes at no point. Suppose in addition that $\overline{f}$ is in $\mathcal{A}$ whenever f is in $\mathcal{A}$. Then the closure of $\mathcal{A}$ is $\mathcal{C}(X, \mathbb{C})$.

17 Special Topics

These are special topics that do not directly relate to any of the texts (Billingsley, Bass, Munkres). However, they are both interesting and useful, and are therefore included in these notes for reference should they appear in a future iteration of this course.

17.1 Brief Intro to Ergodic Theory

(Notes from *Grundlehren der mathematischen Wissenchaften* by Cornfield, Fomin, Sinai, and Sossinskii). This subsection only covers part of the first chapter.

Erdogic theory is the study of motion in a measure space. Particular spaces that are relevant to the study of erdogic theory are (1) the m-dimensional torus with the normalized Haar measure; (2) an m-dimensional compact closed oriented C^∞-class manifold with the differential measure; (3) the space of sequences where each coordinate assumes values from a fixed finite or countable set; (4) the space of real-valued functions. An **automorphism** of a measure space is a bijection T such that $\mu(A) = \mu(TA) = \mu(T^{-1})$. An **endomorphism** is a surjection T of a measure space M onto itself such that $\mu(A) = \mu(T^{-1}A)$.

17.1.1 Liouville's Theorem

Suppose $\{T^t\}$ is a one-parameter group of automorphisms of a measure space and $t \in \mathbb{R}^1$, with $T^{t+s} = T^t(T^s)$. Then $\{T^t\}$ is a **flow** if for any measurable f on M the function $f(T^t x)$ is measurable on $M \times \mathbb{R}^1$. If $\{T^t\}$ is a semigroup, we equivalently define a **semiflow**.

Liouville's Theorem

The measure μ with density p of class C^∞, i.e. the measure with differential $d\mu = p(x)dx_1...dx_m$ is invariant to $\{T^t\}$ if and only if we have

$$\sum_{k=1}^{m} \frac{\partial}{\partial x_k}(pX_k) = 0.$$

17.1.2 Poincaré Recurrence

Suppose T is an endomorphism of $(M, \mathcal{A}, \mu)$ and $A \in \mathcal{A}$. Then $x \in A$ is a **recurrence point** if $T^n(x) = A$ for at least one $n > 0$.

The **Poincaré Recurrence Theorem** suggests that for any endomorphism T and any $A \in \mathcal{A}$ μ-a.e. $x \in A$ is a recurrence point.

17.1.3 Birkhoff-Khinchin Ergodic Theorem

Suppose $(M, \mathcal{A}, \mu)$ is a space with normalized measure and $f \in L^1$ on this space. Then for μ-a.e. $x \in M$, the following limits exist and are equal to each other:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^{-k} x) = \frac{1}{2n+1} \sum_{k=-n}^{n} f(T^k x).$$

17.2 Henstock-Kurzweil Integration

(Notes from *The Kurzeil-Henstock Integral for Undergraduates* by Fonda).

The Henstock-Kurzweil integral is a generalization of the Lebesgue and Riemann integrals. While $\sin(1/x)$ is not Lebesgue integrable and $\mathbf{1}_\mathbb{Q}$ is not Riemann integrable, both functions are Henstock-Kurzweil integrable.

Let P be a **tagged partition** of $[a, b]$; that is, a partition $a = u_0 < ... <$

$u_n = b$ with a tag associated with each interval, $t_i \in [u_{i-1}, u_i]$. Let the Riemann sum for a function f be

$$\sum_P f = \sum_{i=1}^{n} f(t_i)\Delta u_i.$$

Let $\delta : [a, b] \to (0, \infty)$. δ is called a **gauge**; a partition is δ-fine if $[u_{i-1}, u_i] \subset [t_i - \delta(t_i), t_i + \delta(t_i)]$. Then I is the **Henstock-Kurzweil integral** of f if for every $\varepsilon > 0$ there exists a gauge δ such that whenever P is $\delta - fine$,

$$\left|\sum_P f - I\right| < \varepsilon.$$

17.3 Basic Algebraic Topology

(Notes taken from *Algebraic Topology* by Hatcher). Some background in abstract algebra is helpful in understanding this section.

A **path** in a space X is a continuous map $f : I \to X$ where I is the unit interval. We precisely define the idea of continuously deforming a path, keeping its endpoints fixed. A **homotopy** of paths in X is a family $f_t : I \to X$, such that $f_t(0) = x_0$ and $f_t(1) = x_1$ are independent of t; $F : I \times I \to X$ defined by $F(s, t) = f_t(s)$ is continuous. When two paths f_0 and f_1 are connected by such a homotopy, they are called **homotopic**. The equivalence class of a path under the equivalence relation of homotopy is called the **homotopy class of** f, and is denoted $[f]$.

Suppose we restrict our attention to paths with the same starting and ending point. These paths are called **loops**, and the common starting and ending point is called a **basepoint**. The set of all homotopy classes of loops at the same basepoint is denoted $\pi_1(X, x_0)$. This set is a group with respect to $[f \times g]$, and is known as the **fundamental group** of X at x_0.

A set is **path connected** if there exists a path connecting every pair of points, such that the path is contained in the set. X is **simply connected** if there is a *unique* homotopy class of paths connecting any two points in X; in other words, if it is path-connected and has a trivial fundamental group.