1 Introduction

1.1 Catch-up Prerequisites

Basic Topological spaces and analysis. Definitions of basic concepts including continuity, convergence, topological spaces, it would be sensible to know what Riemann integration. Basic linear algebra, what a vector space is and what a norm is. I will briefly mention Hilbert spaces and Banach spaces but not introduce them.

2 Measure Spaces

2.1 What is a measure space?

We begin with a formal definition of what a measure space is. We have a triple $(E, \mathcal{E}, \mu)$ here E is some set. It is the place that we want to 'measure' parts of. It could be $\mathbb{R}^n$ or some space of possible outcomes of an experiment like coin tossing. $\mathcal{E}$ is called the σ-algebra. A σ-algebra is a set of subsets of E. These will be the sets which we can measure. The sets in a σ-algebra must obey a set of rules which are similar to those obeyed by the open or closed sets in a topology.

(i) $\emptyset, E \in \mathcal{E}$,

(ii) $A \in \mathcal{E} \Rightarrow A^c \in \mathcal{E}$.

(ii) $\{A_n, n \in \mathbb{N}\} \subset \mathcal{E} \Rightarrow \bigcup_n A_n \in \mathcal{E}$. An easy way to think about this is that a σ-algebra is closed under taking complements and countable unions or intersections. The very easiest example of a pair $(E, \mathcal{E})$ to think about is if E is some finite set of points for example $E = \{1, 2, 3\}$. and $\mathcal{E}$ is all the possible subsets. We say a subset, $\mathcal{A}$, of $\mathcal{E}$ generates the σ-algebra if $\mathcal{E}$ is the smallest σ-algebra containing $\mathcal{A}$.

The final element μ is the measure. This tells us how big each of the sets are. Fomally μ is a function from $\mathcal{E}$ to $[0, \infty]$. It also has to follow a set of rules.

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

(ii) $\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n)$ when the A_n are all pairwise disjoint.

This last property is called countable additivity.

2.2 Borel σ-algebras

Apart from countable state spaces the most common explicit σ-algebra which appears in applications in the Borel σ-algebra. This is generated when we have a topological space E, and we define the Borel σ-algebra $\mathscr{B}$ to be the smallest σ algebra which contains all the open sets of E.

2.3 π-systems and d-systems

It is useful in several proofs and applications of measure theory to look at other structure on subsets of E. So here we quickly give the definitions.

- A π-system, $\mathcal{A}$, is a set containing $\emptyset$ with the property that if $A, B \in \mathcal{A}$ then $A \cap B \in \mathcal{A}$.
- A d-system, $\mathcal{D}$, is a set containing E with the property that if $A, B \in \mathcal{D}, A \subset B$ then we have $B\ A \in \mathcal{D}$ and that if $A_1 \subset A_2 \subset A_3 \cdots \subset \mathcal{D}$ then $\bigcup_n A_n \in \mathcal{D}$.

A very useful theorem in measure theory is

Theorem 1. *If we have two measures μ_1, μ_2, on a measurable space $(E, \mathcal{E})$ and there exists $\mathcal{A}$, a π-system generating $\mathcal{E}$ on which μ_1 and μ_2 agree then $\mu_1 = \mu_2$.*

2.4 Lebesgue Measure

Lebesgue measure is probably the most famous and fundamental measure. All the details of its construction would take too long. It is a measure on $\mathbb{R}^n$ and is that which corresponds to our intuitive idea of how bit a set is. In 1-D it is defined by

$$\mu((a,b]) = b - a,$$

where $(a,b]$ is any interval. This property completely determines Lebesgue measure on the Borel σ-algebra of $\mathbb{R}$.

2.5 Measure Theoretic Formulation of Probability

In Probability theory it is common to write the measure space as $(\Omega, \mathscr{F}, \mathbb{P})$ with the additional assumption that $\mathbb{P}(\Omega) = 1$. In this setting Ω is the set of all possible individual outcomes (of the experiment, or random process). For an example suppose we are going to toss a coin twice and we want to look at what results we get. Our set is that of all possible sequences

$$\Omega = \{TT, TH, HT, HH\}.$$

We can simply define the σ-algebra to be $\mathcal{P}\Omega$. Then we can see that if we want to see the probability of something occuring, for instance at least one head appearing, then this defines a subset of Ω which is in $\mathscr{F}$ and we can find this probability. $\mathbb{P}($ At least one head appears $) = \mathbb{P}(\{TH, HT, HH\})$.

2.6 Exercises

Here are some hopefully straightforward exercises:

1. Prove that if $(A_n, n \in \mathbb{N}) \subset \mathcal{E} \Rightarrow \bigcap_n A_n \in \mathcal{E}$.
2. Prove that if E is a countable set then $\mathcal{P}E$ is a σ-algebra.
3. Is it always the case that if all the A_n are in $\mathcal{E}$ then $\mu\left(\bigcup_n A_n\right) \le \sum_n \mu(A_n)$.
4. If A_n are all in $\mathcal{E}$ and $\mu\left(\bigcap_n A_n\right) = 0$ is it necessarilly the case that $\bigcap_n A_n = \emptyset$.
5. If $\mathcal{E}$ is both a π-system and a d-system prove that it is a σ-algebra.
6. Speculate on how Lebesgue measure is defined in higher dimensions.

3 Functions

3.1 Measurable Functions

Having our measure spaces allows us to start looking at functions on measure spaces. That is, if $(E_1, \mathcal{E}_1)$ and $(E_2, \mathcal{E}_2)$ are measureable spaces we can look at functions

$$f : E_1 \to E_2.$$

We would like to restric to a set of functions which works with our σ-algebra so we define a *measurable function* to be s.t.

$$A \in \mathcal{E}_2 \Rightarrow f^{-1}(A) \in \mathcal{E}_1.$$

This is analagous to the continuity definition for topological spaces.

There is a useful theorem for finding out the properties of measureable functions.

Theorem 2 (Monotone Class Theorem). *Let $(E, \mathcal{E})$ be a measurable space and let $\mathcal{A}$ be a π-system generating $\mathcal{E}$. Let $\mathcal{V}$ be a vector space of bounded functions $f : E \to \mathbb{R}$ then if*

1. *$1 \in \mathcal{V}$ and $1_A \in \mathcal{V}$ for every $A \in \mathcal{A}$.*
2. *If f_n is a sequence of functions in $\mathcal{V}$ with $f_n \uparrow f$ for some bounded functions f then $f \in \mathcal{V}$.*

Then $\mathcal{V}$ will contain all the bounded measurable functions.

3.2 Random Variables

A random variable is a measurable function from a probability space. For example if we again have the probability space generated by tossing a coin twice. Then if X counts the number of heads, it is a random variable with landing space $\mathbb{N}$ with σ-algebra $\mathcal{P}\mathbb{N}$ often the landing space of a random variable is not made specific. In particular its σ algebra may not be made explicit. The random variable induces a measure on its landing space

$$\mu_x = \mathbb{P} \circ X^{-1}.$$

This is called the law of X.

3.3 Convergence of Measurable Functions.

There are two types of convergence for measurable functions which are different from those for normal functions. There is an additional one for random variables. They are

Convergence almost everywhere this means that if $f_n \to f$ almost everywhere (a.e. or almost surely on a probability space) then the set N on which $f_n(x)$ does not converge to $f(x)$ has $\mu(N) = 0$. In general almost everywhere or almost surely is used to refer to a property that holds everywhere except a set which has measure 0.

Convergence in measure f_n converges to f in probability if $\mu(\{x : |f_n(x) - f(x)| > \epsilon\}) \to 0$ as $n \to \infty$ for each ϵ. This is also called convergence in probability on a probability space.

These two notions are close to being equivalent (at least on finite measure spaces). This is made clear by the following theorem.

Theorem 3. *Let f_n be a sequence of measurable functions on a measure space $(E, \mathcal{E}, \mu)$.*

1. If $\mu(E) < \infty$ (the space is finite) then convergence a.e. implies convergence in measure.

2. If the f_n converge in probability then there exists a subsequence f_{n_k} which converges a.e.

Because probability spaces are always finite we can say convergence a.s. implies convergence in probability.

The last type of convergence is *convergence in distribution.* The distribution function of a random variable is $F_X = \mathbb{P}(X \leq x)$. Then we say $X_n \to X$ in distribution if $F_{X_n}(x) \to F_X(x)$ for every x which is a continuity point of F_X. This differes from the others as the X_n, X do not need to be defined on the same probability space. Again this is related to the other forms of convergence.

Theorem 4. *Let X_n be a sequence of random variable and X another random variable. If they are all defined on the same probability space the convergence in probability implies convergence in distribution. If $X_n \to X$ in distribution then there exist other random variables $\tilde{X}_n, \tilde{X}$ defined on the same probability space such that $X_n \to X$ a.s.*

3.4 Exercises

Exercises on functions:

1. Prove that it is only necessary to check the measurablity criterion on a π-system generating the σ-algebra.

2. Prove that a continuous function between two topological spaces with their Borel σ-algebras is measurable.

3. Prove that if f, g are measureable functions into $\mathbb{R}$ with its Borel σ-algebra then fg and $f + g$ are also measurable.

4. Prove that if f_n are all measurable functions into $\mathbb{R}$ with its Borel σ-algebra then $\inf_n f_n$, $\sup_n f_n$, $\liminf_n f_n$ and $\limsup_n f_n$ are all measurable.

5. Try and come up with sequences that converge either in measure or probability but not in both.

4 Integration

The theory of Lebesgue integration allows you to integrate functions with respect to a measure. This is written as $\int_E f \mathrm{d}\mu$, $\mu(f)$ or $\int_E f(x)\mu(\mathrm{d}x)$. I wont go into the detail of how the integral is constructed or shown to exist as it is not needed for most applications. The notaion μ-integrable is used of a function when $\mu(|f|)$ exists and is finite.

4.1 Convergence of Integrals

There are three main convergence theorems for integrals.

Theorem 5 (Monotone Convergence)**.** *Let f_n be a sequence of positive functions such that $f_n \uparrow f$ then $\mu(f_n) \uparrow \mu(f)$.*

Theorem 6 (Fatou's Lemma)**.** *Let f_n be a sequence of positive functions then $\mu(\liminf_n f_n) \leq \liminf_n \mu(f_n)$.*

Theorem 7. *Let f_n be a sequence of measurable functions and $f_n(x) \to f(x)$ and that $|f(x)| \leq g(x)$ for every x and $\mu(g) < \infty$. Then f, f_n are integrable and $\mu(f_n) \to \mu(f)$. We call g the dominating function.*

4.2 Product Measure and Fubini's Theorem

If $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$ are probability space then we can construct the probability space $(E_1 \times E_2, \mathcal{E}_1 \times \mathcal{E}_2, \mu_1 \times \mu_2)$. We do this by looking at the π-system of sets of the form $A \times B$ where $A \in \mathcal{E}_1$ and $B \in \mathcal{E}_2$. Then we can define $\mathcal{E}_1 \times \mathcal{E}_2$ to be the σ-algebra generated by this π-system. We can also extend a measure defined on this π system by $\mu_1 \times \mu_2(A \times B) = \mu_1(A)\mu_2(B)$. Fubini's theorem says that if $f : E_1 \times E_2$ is a $\mathcal{E}_1 \times \mathcal{E}_2$ measurable function then

$$\mu_1 \times \mu_2(f) = \int_{E_1} \left(\int_{E_2} f(x_1, x_2)\mu_2(\mathrm{d}x_2) \right) \mu_1(\mathrm{d}x_1).$$

4.3 Lebesgue Spaces

An important class of function spaces are Lebesgue spaces of L^p spaces. Essentially for $1 \leq p < \infty$ these are spaces of functions from E to $\mathbb{R}$ such that $\mu(|f|^p) < \infty$ equipped with the norm $\|f\|_p = \mu(|f|^p)^{1/p}$. However this is not a true norm since there are many functions with $f = 0$ a.e. therefore we have to consider instead of f the equivalence class of functions $\{g : g = f a.e.\}$. This complication is often suppressed and you don't really notice it. L^∞ is the space of functions whose essential supremum is finite. The essential supremum is defined by

$$\|f\|_\infty = \inf_{N \in \mathcal{N}} \sup_{x \in E - N} |f(x)|.$$

Where $\mathcal{N}$ is the collection of sets in $\mathcal{E}$ where $\mu(N) = 0$ (null sets). The L^p spaces are Banach spaces with this norm and L^2 is a Hilbert space with the inner product

$$< f, g >= \int_E fg \mathrm{d}\mu.$$

4.4 Exercises

1. Find a sequence of measurable functions f_n such that f_n converges to some function f with $\mu(|f|) < \infty$, but $\mu(f_n)$ doesn't converge to $\mu(f)$.

2. The sequence

$$f_n(x) = \sum_{k=1}^{2^n} \sqrt{\frac{k}{2^n}} 1_{[(k-1)/2^n, k/2^n)}$$

converges to $f(x)$. Prove that $\mu(f_n)$ converges to $\mu(f)$ where μ is Lebesgue measure on $[0,1]$.

3. Check that the π-system defined in the making of the product measure space is in fact a π-system.
4. Find a function which is in $L^2(\mathbb{R})$ but not in $L^1(\mathbb{R})$ and vice versa.

5 Inequalities

In both probability and PDEs common inequalities appear very frequently and may not always be referenced directly. Here is a list of what I think are the most common and useful.

Theorem 8 (Markov's Inequality). *This is sometimes also called Chebychev's inequality but that is also a name given to a corollary of Markov's inequality. This says that if X is a random variable and $\lambda > 0$ then*

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}.$$

Theorem 9 (Jensen's Inequality). *Suppose we have ϕ is a convex function (for $t \in (0,1), \phi(tx+(1-t)y) \leq t\phi(x)+(1-t)\phi(y)$) and $(E, \mathcal{E}, \mu)$, a measure space with $\mu(E) < \infty$. Then we will have*

$$\mu(\phi(f)) \leq \phi(\mu(f)).$$

Theorem 10 (Holder's Inequality). *If we have $1/p + 1/q = 1$ and $f \in L^p, g \in L^q$ then we have*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

In the case where $p = q = 2$ this is called Cauchy-Schwarz inequality.

Theorem 11 (Minkowski's Inequality). *This one is easy to remember because it is the triangle inequality for L^p norms, but it is not so easy to prove. If f, g are in L^p then*

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

Theorem 12 (Young's Inequality). *There is a more complicated inequality of young involving convolutions. This inequality is not about measure theory but it is very useful particularly in PDEs after using Cauchy-Schwarz. If a, b are real numbers*

$$|ab| \leq \frac{1}{2}(a^2 + b^2),$$

and following this

$$|ab| \leq \frac{\epsilon}{2}a^2 + \frac{1}{2\epsilon}b^2.$$

5.1 Exercises

1. Show if X is a Normal random variable with mean 0 and variance 1. Show that

$$\mathbb{P}(X \geq \alpha) \leq \frac{1}{2\alpha^2}.$$

2. Show that if $f, g \in L^2(\mu)$ then

$$\left|\int f(x)g(x)\mu(\mathrm{d}x)\right| \leq \frac{1}{2}\int |f(x)|^2\mu(\mathrm{d}x) + \frac{1}{2}\int |g(x)|^2\mu(\mathrm{d}x).$$

3. Show that if $f \in L^1 \cap L^2$ then

$$\int_0^T f(x)^2 \mathrm{d}x \leq \frac{1}{T}\left(\int_0^T f(x)\mathrm{d}x\right)^2.$$

6 Useful Things I have skipped

These are mainly related to probability and can be found in the probability and measure notes.

1. Borel-Cantelli Lemmas,
2. Kolmogorov 0-1 Law,
3. Fourier Theory Relating to L^2.
4. Uniform Integrability