Ergodic Theory - Jan-Mar. 2011

(Measure Theory Handout)

January 7, 2011

The first six sections of these notes are a reminder of the key results in measure theory. The remaining sections are more details on their proofs, for the benefit of enthusiasts.

1 Introduction

Probably the first question to answer is the following.

Question 1.1. *What is a measure?*

It is better to start from a concrete perspective and then consider the more abstract point of view.

Example 1.2 (Concrete Example)**.** *Given , say, the unit interval* $[0, 1]$ *we can naturally describe the size of a subinterval* $[a, b]$ *(or* $[a, b)$ *or* $(b, a]$ *or* (b, a) *) by its length* $b-a$*. Similarly, for any finite (disjoint) union of subintervals* $\mathcal{A}$ *we can add up their individual lengths. However, as the sets get more complicated, it gets harder to say exactly what their "measure" is. Let* $\mathcal{B}$ *be the collection of sets we get by all "reasonable" operations on intervals (countable intersections, countable unions, differences etc.)*

More generally, let X be any set and let $\mathcal{B}$ be a collection of subsets $\mathcal{B}$ of X. We call $\mathcal{B}$ a sigma algebra if:

1. $X, \emptyset \in \mathcal{B}$
2. $B \in \mathcal{B} \implies X - B \in \mathcal{B}$
3. $\{B_n\}_{n=1}^{\infty} \implies \cup_{n=1}^{\infty} B_n \in \mathcal{B}$

$\mathcal{B}$ gives the collection of sets whose measure we want to define. Intuitively, the measure of a set $B \in \mathcal{B}$ is its "size". More formally,

Definition 1.3. *A measure* m *is a positive function* $m : \mathcal{B} \to \mathbb{R}^+$ *such that*

1. $m(\emptyset) = 0$*;*
2. $m(\cup_n B_n) = \sum_n m(B_n)$*, whenever* $B_n \in \mathcal{B}$ *are disjoint.*

Definition 1.4 (Classic Example: Lebesgue measure)**.** *For* $X = [0, 1]$ *and* $\mathcal{B}$ *coming from countable unions and intersections, etc., of intervals, the "usual measure" which gives intervals their lengths is Lebesgue (or Haar) measure.*

Of course, not all measures look like Lebesgue measure:

Definition 1.5 (Example: Dirac measures)**.** *We could fix any point* $x_0 \in X$ *could let* $\mu : \mathcal{B} \to \mathbb{R}$ *be defined by*

$$\mu(B) = \begin{cases} 1 & \text{if } x_0 \in B \\ 0 & \text{if } x_0 \notin B \end{cases}$$

Here every set has measure one (if it contains x_0*) or zero (if it doesn't contain* x_0*).*

Definition 1.6 (Geometric examples)**.** *For any manifold* M *with a Riemannian metric we can associate a natural measure (the Riemannian volume).*

In the particular case that $m(X) < +\infty$ we call m a finite measure. In the particular case that $m(X) = 1$ we call m a probability measure (as in these examples).

2 Kolmogorov's theorem and Borel sigma algebras

Usually, we know what we want the measure to be on simpler sets (e.g., subintervals of $[0, 1]$). But these don't necessarily form a sigma algebra. We can let $\mathcal{B}$ be the smallest sigma algebra containing these sets (by adding in all countable unions and intersections, etc.), but ...

Question 2.1. *How do we know that the measures* $m(B)$ *are defined for all* $B \in \mathcal{B}$*?*

This is solved by the following extremely useful theorem:

Theorem 2.2 (Kolmogorov extension theorem)**.** *Let* $\mathcal{A}$ *be a collection of sets closed under finitely many unions and intersections. Let* $\mathcal{B}$ *be the sigma algebra given by countable intersections and unions of sets in* $\mathcal{A}$*. Then* $m : \mathcal{A} \to \mathcal{R}$ *always extends to a unique measure* $m : \mathcal{B} \to \mathbb{R}$*.*

The proof appears in Appendix 9

Equivalently, we could say that $\mathcal{B}$ is the smallest sigma algebra containing $\mathcal{A}$.

The definition of a sigma algebra looks similar to that of a topology (i.e., a collection of open sets) except for condition 2. By far, the most common sigma algebra in Dynamical Systems is the following:

Example 2.3 (Borel sigma algebra)**.** *If we assume that* X *is a metric space and* $\mathcal{A}$ *is the collection of all open sets (i.e., the topology) and* $\mathcal{B}$ *is the smallest sigma algebra containing* $\mathcal{A}$*, then* $\mathcal{B}$ *is called the Borel sigma algebra. We call a set* $B \in \mathcal{B}$ *a Borel set.*

Question 2.4. *Why bother with sigma algebras? Can we define the measure of any set* $B \subset X$*?*

Unfortunately not. There exist lots of sets which are not in the Borel sigma algebra and thus we don't know how to define their measures. However, things are even worse: one can "construct" examples of sets to which Lebesgue measure cannot be extended. This requires the Axiom of Choice and is explained in section 8.

Remark 2.5 (Remark on other sigma algebras). In the most general setting, we are free to fix whatever sigma algebra we like. For example, we could take $\mathcal{B}$ to consist of just two sets $\{X, \emptyset\}$ (called the trivial sigma algebra) or, at the other extreme, the sigma algebra $\mathcal{B} = \{B \subset X\}$ consisting of all subsets.

3 Measurable functions and Integration

In the theory of Riemann integrable functions $f : [0,1] \to \mathbb{R}$ one approximates the function by functions constant on intervals - and hope that it converges.

Question 3.1. *What are sensible function in measure theory?*

Let X have a sigma algebra $\mathcal{B}$.

Definition 3.2. *A function $f : X \to \mathbb{R}$ is called measurable if for any (Borel) measurable set $A \subset \mathbb{R}$ the set $f^{-1}(A) \in \mathcal{B}$.*

Example 3.3. *If $B \subset X$ is in $\mathcal{B}$ then we can define a measurable function by*

$$\chi_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{otherwise} \end{cases}$$

which we call the indicator function *(or* characteristic function*).*

Example 3.4. *If X is a compact metric space and $\mathcal{B}$ is the Borel sigma algebra then any continuous function $f : X \to \mathbb{R}$ is measurable.*

In the trivial case that $f : X \to \mathbb{R}^+$ is a "step functions" $f(x) = \sum_{i=1}^n a_i \chi_{B_i}(x)$ ($a_i \in \mathbb{R}$ and $B_i \in \mathcal{B}$) we can define its integral

$$\int f(x)dm(x) = \int \left(\sum_{i=1}^n a_i \chi_{B_i}(x) \right) dm(x) := \sum_{i=1}^n a_i m(B_i)$$

More generally, for any measurable function $f : X \to \mathbb{R}^+$ we can define its (Lebesgue) integral by approximation:

$$\int f(x)dm(x) = \sup \left\{ \int \left(\sum_{i=1}^n a_i \chi_{B_i}(x) \right) dm(x) \,:\, 0 \le \sum_{i=1}^n a_i \chi_{B_i}(x) \le f(x) \right\}$$

When this quantity is finite the function is called (Lebesgue) integrable.

Example 3.5. *Let*

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is irrational} \\ 0 & \text{if } x \text{ is rational.} \end{cases}$$

This is not Riemann integrable, but it is Lebesgue integrable with $\int f(x)dm(x) = 1$. (We can simply choose the sets B_i not to include any rational numbers, which have zero measure.)

Assume that we have two measures m_1 and m_2. We say that m_1 is *absolutely continuous* with respect to m_2 if for any set $B \in \mathcal{B}$ with $m_1(B) = 0$ we automatically have $m_2(B) = 0$. This is usually denoted by $m_2 << m_1$.

Example 3.6. *Given m_2 and $f \in L^1(X, m_1)$ we can define m_1 by $m_1(A) = \int_A f(x)dm_2(x)$, for all $A \in \mathcal{B}$. We see that m_1 is absolutely continuous with respect to m_2*

In fact the converse is also true:

Theorem 3.7 (Radon-Nikodym Theorem)**.** *If $m_2 << m_1$ then there exists $f \in L^1(X, m_1)$ such that $m_1(A) = \int_A f(x)dm_2(x)$, for all $A \in \mathcal{B}$.*

The proof is explained in section 12

Usually, we write $f(x) = \frac{dm_2}{dm_1}$, and call this the Radon-Nikodym derivative.

4 "Almost everywhere" and "almost all"

In measure theory we are basically interested in sets of non-zero measure. Any set B which has zero measure is essentially "invisible" to the measure (and us ...).

The most important terminology in ergodic theory is the most basic:

Definition 4.1. *We say that a property holds for almost all points $x \in X$ if there exists a set $N \in \mathcal{B}$ such that the Property holds for all $x \in X - N$ (i.e., it holds for all points not in the negligible set N).*

We can also say that the property holds "almost everywhere".

Definition 4.2. *Usually we abbreviate this to "a.e." and "a.a. $x \in X$". Sometimes it is just implied (or forgotten! - but in any case not written).*

Of course N may or may not seem quite large or small in a topological sense, depending on what the measure m is:

Example 4.3 (Lebesgue measure)**.** *A single point, or even a countable set of points, always has zero Lebesgue measure. The usual middle third Cantor set also only has Lebesgue zero measure.*

For Lebesgue measure is doesn't make much sense to talk about properties of individual points - since they have zero measure.

Example 4.4 (Dirac measure)**.** *On the other hand, if we are using a Dirac measure then a set B has zero measure precisely when it doesn't contain the point. In this case, a.e. results are equivalent to properties of the point x.*

This distinction becomes rather important when we consider spaces of functions $f : X \to \mathbb{R}$.

4.1 Spaces of Functions

We can consider the space of all measurable functions $f : X \to \mathbb{R}$ such that $\int |f(x)|dm(x) < +\infty$. *However, we want to identify functions which differ on a set of zero measure* (i.e., functions that are the same, almost everywhere). We denote the resulting class of functions as $L^1(X, m)$. This is easily seen to be a vector space. Moreover, it has a norm $||f|| = \int |f(x)|dm(x)$, with respect to which it is a Banach space.

Similarly, we write $L^2(X, m)$ as the space of measurable functions $f : X \to \mathbb{R}$ (again identifying functions which differ on a set of zero measure) such that $\int |f|^2 dm(x) < +\infty$, etc. This is a Hilbert space.

Example 4.5 (Lebesgue measure)**.** *Here $L^1([0, 1], m)$ is rather a large space. However, the function we considered before:*

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is irrational} \\ 0 & \text{if } x \text{ is rational.} \end{cases}$$

we can consider to be the same as the constant function 1*, since they differ on a set of zero measure (namely, the rational numbers on the unit interval).*

Example 4.6 (Dirac measure)**.** *This space is only one dimensional: Any two functions which take the same value at x_0 are identified, since the rest of the space has measure zero!*

5 Limits and integrals

We need to know about how sequences of functions in $L^1(X, m)$, and the sequence of their integrals, are related.

Interchanging limits (and integrals) is always a tricky problem in any branch of analysis - and measure theory is no different. Let $(f_n)_{n=0}^{\infty}$ be a sequence of integrable functions.

A basic question is:

Question 5.1. *If $f_n(x)$ converges to $f(x)$ a.e. then when can we write $\int f(x)dm(x) = \lim_{n\to\infty} \int f_n(x)dm(x)$?*

We need to make some extra assumptions, as the following simple example shows.

Example 5.2. *For $n \geq 0$*

$$f_n(x) = \begin{cases} 2^n & \text{if } 0 \leq x \leq \frac{1}{2^n} \\ 0 & \text{otherwise} \end{cases}$$

By construction, we see that $\int f_n dm = 1$, for all $n \geq 0$. However, $\lim_{n\to\infty} f_n = 0$, a.e.

However, if we impose condition on the sequence we can get convergence:

Theorem 5.3 (Convergence Theorems)**.** *Assume that $f_n(x)$ converges to $f(x)$ a.e.. Assume that either*

1. *f_n are monotone (i.e., $f_n(x) \leq f_{n+1}(x)$, a.e.); or*
2. *f_n are dominated (i.e., there exists $g \in L^1(X, m)$ with $|f_n(x)| \leq g(x)$, a.e.,*

then $\int f(x)dm(x) = \lim_{n\to\infty} \int f_n(x)dm(x)$.

On the other hand, if we don't choose to make any other assumptions, we can still get the following inequality:

Theorem 5.4. *Fatou's Lemma*

$$\int \left(\liminf_{n\to\infty} f_n \right) d\mu \leq \liminf_{n\to\infty} \int f_n d\mu.$$

The proofs of these results are outlined in section 13.

We might also reasonably ask about the connection between convergence of functions $f_n(x)$ a.e., and convergence in the norm on $L^1(X, m)$.

Example 5.5 (Convergence a.e. doesn't imply convergence in $L^1(X, m)$))**.** *The previous example shows this. We had a sequence of functions $\{f_n\}$ which converges a.e. to the zero function $f = 0$. Unfortunately, they don't converge in $L^1(X)$ since*

$$||f_n - 0||_1 = \int |f_n(x) - 0| dx = \int |f_n(x)| dx = 1 \not\to 0$$

Unfortunately, the other implications doesn't hold either. We have essentially already seen the following:

Example 5.6 (Convergence in $L^1(X, m)$ doesn't imply convergence a.e.)**.** *For $2^k + i \leq n \leq 2^k + (i + 1)$, with $0 \leq i \leq 2^k - 1$ and $k \geq 0$ we can define*

$$f(x) = \begin{cases} 1 & \text{if } x \in [\frac{i}{2^k}, \frac{i+1}{2^k}] \\ 0 & \text{otherwise} \end{cases}$$

In this case we see that $||f||_1 = \int |f_n(x)| dx = 2^{-k} \to 0$, i.e., f_n converges to the zero function in the $|| \cdot ||_1$ norm. However, for every point $x \in [0, 1]$ we see that $f_n(x) \not\to 0$.

6 Probability measures on compact metric space

In most cases of interest to us, X will be a compact metric space and $\mathcal{B}$ the Borel Sigma algebra. In this context, we have much more information on the set of possible probability measures.

Let $C(X, \mathbb{R})$ be the vector space of continuous functions $f : X \to \mathbb{R}$. We can define a norm on $C(X, \mathbb{R})$ by $||f||_\infty = \sup_{x \in X} |f(x)|$. Given any probability measure m we can associate a continuous linear map $L_m : C(X, \mathbb{R}) \to \mathbb{R}$ by

$$L_m(f) = \int f(x) dm(x).$$

It is easy to see that:

1. L_m is continuous;
2. L is positive, i.e., if $f \geq 0$ then $L(f) \geq 0$; and
3. if $\underline{1}(x) = 1$ is the function that takes the constant value 1, then $L(\underline{1}) = 1$.

The interesting thing is that the converse is true:

Theorem 6.1 (Riesz representation Theorem)**.** *Let $L : C(X, \mathbb{R}) \to \mathbb{R}$ be a continuous linear functional satisfying (1),(2) and (3) above. Then there exists a measure m such that $L(f) = \int f dm$.*

The proof is described in section 10.

In particular, we can identify the space of probability measures with the space of (certain) linear functionals on the Banach space $C(X, \mathbb{R})$. This has the advantage that we can make use of useful results and ideas from functional analysis. For example, we say that a sequence of probability measures $\{m_n\}_{n=1}^\infty$ converges to a probability measure m in the *weak star topology* if for every continuous function $f : X \to \mathbb{R}$ we have that $\lim_{n \to +\infty} \int f dm_n = \int f dm$.

Theorem 6.2 (Alaoglu's Theorem)**.** *The space of probability measures is compact in the weak star topology (i.e., Given a sequence $\{m_n\}_{n=1}^\infty$ there exists a probability measure m and a subsequence $\{m_{n_i}\}$ which converges to m in the weak star topology).*

The proof is described in section 11.

Finally, there are a couple of amusing results that tell us that measurable setting is not so far from the continuous setting as one might expect.

Theorem 6.3 (Ergorof Theorem)**.** *Let $\{f_n\}_{n=1}^\infty$ we a sequence of measurable functions. Let f be a measurable function and assume that $\lim_{n \to +\infty} f_n(x) = f(x)$, almost everywhere. Then there exists a set $Y \in \mathcal{B}$ such that $m(X - Y) < \epsilon$ and $f_n \to f$ uniformly on Y.*

The second result is the following.

Theorem 6.4 (Lusin's Theorem)**.** *Given a measurable function $f : X \to \mathbb{R}$ and $\epsilon > 0$ we can find a set $Y \in \mathcal{B}$ such that $m(X - Y) < \epsilon$ and the restriction $f : Y \to \mathbb{R}$ is continuous.*

The proofs of these results are described in section 14.

7 The cast list

Lebesgue (1875-1941) helped to found the main ideas in analysis. Kolmogorov (1903-1987) went on from being a train conductor to founding the modern theory of probability theorey (circa 1933). Radon (1887-1956) proved the theorem for $\mathbb{R}^n$ in 1913 and Nikodym (1887-1974) the general case in 1930. Banach (1892-1945) establish much of the analysis of vector spaces, often working at the "scottish cafe" in Warsaw. Lusin (1883-1950) continued in Mathematics, only after some persuasion from his supervisor Ergorof. Tragically, Ergorof (1869-1931) was arrested and died after a hunger strike. Dirac (1902-1984) was a Bristol born physicist who won the Nobel prize in 1933.

8 Proof of existence of non-measurable sets

We can partition he unit interval $[0,1]$ it into equivalence classes by the equivalence relation: $x \sim y$ *if and only if* $x - y$ *is a rational number.* Clearly, the equivalence classes are disjoint, nonempty sets. Let V be a set containing exactly one representative from each equivalence class (constructed using the Axiom of Choice). We claim that V is non-measurable (i.e., Lebesgue measure is not defined for these sets).

Lemma 8.1. *If q and q' are two distinct rationals, and $V + q$ is the translation of V by q to each element of V, then $V + q$ and $V + q'$ are disjoint.*

Proof. Indeed, if any element x lies in both $V + q$ and $V + q'$, then it must equal $y + q$ and $z + q'$, say, with $y, q \in V$. But then $y + q = z + q'$, so $y - z = q' - q$ is a rational number. In particular, since V contains exactly one representative from each equivalence class, and $y \sim z$, we deduce that $y = z$. But then $q = q'$, a contradiction. So all the $V + q$ are pairwise disjoint, as claimed. □

Consider the union of all $V + q$, where $q \in \mathbb{Q} \cap [-1, 1]$. Assume for a contradiction that V is measurable.

Lemma 8.2. *V has zero Lebesgue measure.*

Proof. If V is measurable, then so is $V + q$ for all q. Moreover, they all have the same measure, since Lebesgue measure is invariant under translation. Note that the union of all such $V + q$ is contained in $[-1, 2]$, because $V \subset [0, 1]$ and $-1 \leq q \leq 1$. Thus, the Lebesgue measure of the union of the $V + q$'s must be finite, since the union is contained in $[-1, 2]$. Since they are all disjoint, and Lebesgue measure is countably additive, it must be that V has zero Lebesgue measure, as required. □

Now we claim that the union of all the $V + q$, $q \in \mathbb{Q} \cap [-1, 1]$, contains the interval $[0, 1]$. Any $-1 \leq x \leq$ is equivalent to some $y \in V$. In particular, $x - y$ is a rational. Since both x and y are both in $[0, 1]$, $x - y$ is equal to a rational, $q \in [-1, 1] \in \mathbb{Q}$. Therefore, $x = y + q$, so $x \in V + q$.

Therefore, since $[0, 1]$ has measure 1 and is contained in the union of all the $V + q$, if V is measurable then the measure of the union of all the $V + q$ must be at least 1. But we already deduced that the measure of V (and hence of all $V + q$) is 0, so the measure of the union of all the $V + q$ would be zero.

This contradiction shows that V is not a measurable set.

9 Proof of the Kolmogorov Extension Theorem

In order to extend μ from the algebra $\mathcal{A}$ to the sigma algebra $\mathcal{B}$ we first try something too optimistic: For any $A \subset X$ we define

$$\mu^*(A) = \inf \left\{ \sum_n \mu(A_n) : \cup_n A_n \supset A, A_n \in \mathcal{A} \right\}$$

The first claim is easy to check from the definitions:

Lemma 9.1. *μ^* satisfies the following*

1. *$\mu^*(\emptyset) = 0$;*
2. *If $A \subset B$ then $\mu^*(A) \leq \mu^*(B)$; and*
3. *$m(\cup_n B_n) \leq \sum_n m(B_n)$, whenever $B_n \in \mathcal{B}$ are disjoint.*

We then say that μ^* is an outer measure.

The outer measure μ^* is now defined for *any* set $E \subset X$. However, need to restrict a little to get the necessary properties for a measure.

Definition 9.2. *We say that a set E is μ^*-measurable if*

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap (X - E))$$

for any $A \subset X$.

We can introduce a sigma algebra $\mathcal{M} \supset \mathcal{A}$ as follows:

Lemma 9.3. *We have that:*

1. *The collection of μ^*-measurable sets $\mathcal{M}$ is a sigma algebra; and*
2. *$\mu^* : \mathcal{B} \to \mathbb{R}$ is a measure.*

Proof. 1. We need only check the three conditions for a sigma algebra.
(i) Clearly, $\emptyset, X \in \mathcal{M}$
(ii) Suppose E measurable, then $X - E$ is measurable by the symmetry of (9.2).
(iii) Suppose that $E, F \in \mathcal{M}$. For any $A \subset X$ then by (9.2)

$$\begin{aligned}
\mu^*(A) &= \mu^*(A \cap (E \cup F)) + \mu^*(A \cap [X - (E \cup F)]) \\
&= \mu^*(A \cap (E \cup F)) + \mu^*(A \cap ([X - E] \cup [X - F])) \\
&\leq \mu^*(A \cap (E \cup F)) + \mu^*(A \cap (E \cup [X - F])) + \mu^*(A \cap ([X - E] \cup F)) \\
&\quad + \mu^*(A \cap ([X - E] \cup [X - F])) \\
&= \mu^*(A \cap F) + \mu^*(A \cap [X - F]) = \mu^*(A),
\end{aligned}$$

by using (*) twice. In particular, we see from the first line that $E \cup F \in \mathcal{M}$.

We proceed to finite unions. Assume that $E_n \in \mathcal{M}$ are disjoint sets. By induction we can see that if $\cup_{n=1}^{N-1} E_n \in \mathcal{M}$ then so is $\cup_{n=1}^{N} E_n$ since for any $E \subset X$,

$$\begin{aligned}\mu^*(A \cap (\cup_{n=1}^N E_n)) &= \mu^*(A \cap (\cup_{n=1}^{N-1} E_n) \cap E_N) + \mu^*(A \cap (\cup_{n=1}^{N-1} E_n) \cap [X - E_N]) \\ &= \mu^*(A \cap (\cup_{n=1}^{N-1} E_n)) + \mu^*((A \cap E_N)\end{aligned}$$

Finally, we proceed to countable unions. We see that for every $N \geq 1$:

$$\sum_{n=1}^N \mu^*(A \cap E_n) = \mu^*(\cup_{n=1}^N (A \cap E_n)) \leq \mu^*(\cup_{n=1}^\infty (A \cap E_n)) \leq \sum_{n=1}^\infty \mu^*(A \cap E_n)$$

and so letting $N \to +\infty$ gives

$$\lim_{N \to +\infty} \sum_{n=1}^N \mu^*(A \cap E_n) = \mu^*(\cup_{n=1}^\infty (A \cap E_n)) = \sum_{n=1}^\infty \mu^*(A \cap E_n).$$

To see that $\cup_{n=1}^\infty E_n \in \mathcal{B}$ (and thus that $\mathcal{B}$ really is a sigma algebra) observe that for any $N \geq 1$,

$$\begin{aligned}\mu^*(A) &= \mu^*(A \cap (\cup_{n=1}^N (A \cap E_n))) + \mu^*(A \cap (X - \cup_{n=1}^N (A \cap E_n)) \\ &\geq \sum_{n=1}^N \mu^*(A \cap E_n)) + \mu^*(A \cap (X - \cup_{n=1}^N (A \cap E_n))\end{aligned}$$

and the lettiing $N \to +\infty$ gives

$$\mu^*(A) \geq \sum_{n=1}^\infty \mu^*(A \cap E_n)) + \mu^*(A \cap (X - \cup_{n=1}^N (A \cap E_n)).$$

(b) It only remains to observe that μ^* is sigma additive, by taking $A = X$ in (1). □ □

Of course, since $\mathcal{B} \supset \mathcal{A}$ is the *smallest* sigma algebra containing $\mathcal{A}$ the extension to $\mathcal{M}$ automatically gives the extension to $\mathcal{B} \subset \mathcal{M}$.

Remark 9.4. The measure μ^* is complete, i.e., whenever E is μ^*-measurable with $\mu^*(E) = 0$, and $F \subset E$, then F is μ^*-measurable (with $\mu^*(F) = 0$).

We only need to show that if $\mu^*(E) = 0$ then $E \subset X$ is measurable. Let $A \subset X$. By (3)

$$\begin{aligned}\mu^*(A) &\leq \mu^*(A \cap E) + \mu^*(A \cap [X - E]) \\ &\leq \mu^*(A \cap [X - E]) \leq \mu^*(A).\end{aligned}$$

Thus E is measurable.

Lemma 9.5. *The extension μ^* to $\mathcal{M}$ is unique.*

Proof. Assume that ν^* and μ^* are two extensions of μ. Let $A \in \mathcal{M}$ and let $\{A_n\} \subset \mathcal{A}$ with $A \subset \cup_n A_n$. Thus

$$\nu^*(A) \leq \nu^*(\cup_n A_n) \leq \sum_n \nu^*(A_n) = \sum_n \mu(A_n)$$

and so $\nu^*(A) \leq \mu^*(A)$ for every μ^* measurable set. Since

$$\begin{aligned}\mu^*(A) + \mu^*(X - A) &= \nu^*(A) + \nu^*(X - A) \\ &\leq \mu^*(A) + \mu^*(X - A)\end{aligned}$$

□

10 Proof of Riesz Representation Theorem

Given a positive linear functional $u : C^0(X, \mathbb{R}) \to \mathbb{R}$, we define:

$$\mu(U) = \sup\{u(f) : 0 \leq f(x) \leq 1 \text{ and } f(x) = 0 \text{ if } x \notin U\} \text{ if } U \text{ is open; and}$$
$$\mu^*(E) = \inf\{\mu(U) : U \supset E, \text{ where } U \text{ is open }\} \text{ if } E \subset X$$

Here μ^* is defined for *any* set $E \subset X$.

Lemma 10.1. *μ^* is an outer measure, i.e.,*

1. *$\mu^*(\emptyset) = 0$;*
2. *If $A \subset B$ then $\mu^*(A) \leq \mu^*(B)$; and*
3. *$m(\cup_n B_n) \leq \sum_n m(B_n)$, whenever $B_n \in \mathcal{B}$ are disjoint.*

Proof. Parts (1) and (2) are easy, so we need to show (3). Suppose first that the B_n are open, then so is $B = \cup_n B_n$. If $f\ inC^0(X)$ is zero outside of B then its support will be covered by finitely many sets, i.e., $B_{i_1}, \ldots, B_{i_k}$, say. Furthermore, we can choose a "partitions of unity" $f_{i_1}, \ldots, f_{i_k} \in C^0(X)$ supported in the corresponding sets such that $\sum_{j=1}^k f_{i_j}(x) = 1$. Thus $\sum_{j=1}^k f_{i_j}(x) f(x) = f(x)$ and so

$$u(f) = \sum_{j=1}^k u(f_{i_j}(x) f(x))$$

which implies

$$\mu^*(A) \leq \sum_{j=1}^k \mu^*(A_i) \leq \sum_{n=1}^\infty \mu^*(A_n)$$

since $0 \leq f_{i_j}(x) f(x) \leq 1$ and $f_i f$ is supported in A_i.

More generally, if A_i is not open, we can choose open sets $B_i \supset A_i$ and then

$$\mu^*\left(\cup_{n=1}^\infty A_n\right) \leq \mu^*\left(\cup_{n=1}^\infty B_n\right) \leq \sum_{n=1}^\infty \mu^*(B_n)$$

Finally, taking the infimum over open sets B_n proves (3) □

Now it only remains to show that every open set is in the sigma algebra $\mathcal{M}$ of μ^*-measurable functions, then the Borel sigma algebra must be contained $\mathcal{M}$ (by Kolmogorov's Theorem). Since $\mu(U) = \mu^*(U)$ for every open set, μ extends as μ^*, as required.

Lemma 10.2. *Every open set is μ^*-measurable*

Proof. Let $U \subset X$ be open. First assume that $A \subset X$ is open, then $A \cap U$ is also open. Given $\epsilon > 0$ we can choose a continuous function $f : X \to [0, 1]$ whose support $\text{supp}(f) \subset A \cap U$ and

$$\mu^*(A \cap U) = \mu(A \cap U) \leq u(f) + \epsilon$$

Moreover, since $A - \text{supp}(f)$ is open we can also choose a continuous function $g : X \to [0, 1]$ whose support $\text{supp}(g) \subset A - \text{supp}(f)$ and

$$\mu^*(A - \text{supp}(f)) = \mu(A - \text{supp}(f)) \leq u(g) + \epsilon.$$

In particular,

$$\begin{aligned}\mu(A) &\geq u(f+g) = u(f) + u(g) \\ &> \mu^*(A \cap U) + \mu^*(A \cap [X-U]) - 2\epsilon \\ &\mu^*(A) - 2\epsilon\end{aligned}$$

and since ϵ can be arbitrarily small

$$\mu(A) = \mu^*(A) = \mu^*(A \cap U) + \mu^*(A \cap [X-U]).$$

More generally, for any $E \subset X$ and $\epsilon > 0$ choose an open set $A \subset X$ with $\mu^*(E) > \mu^*(A) - \epsilon$. Thus

$$\begin{aligned}\mu^*(E) &\geq \mu^*(A \cap U) + \mu^*(A \cap [X-U]) - \epsilon \\ &\geq \mu^*(E \cap U) + \mu^*(E \cap [X-U]) - \epsilon\end{aligned}$$

and since $\epsilon > 0$ is arbitrary we deduce that $\mu^*(U) = \mu^*(E \cap U) + \mu^*(E \cap [X-U])$, i.e., U is μ^*-measurable. □

11 Proof of Alaoglu's Theorem

Choose a countable sequence $\{f_i\}$ of functions that forms a dense subset of $C^0(X, \mathbb{R})$. This is possible since X is compact. Define $\mu(f) := \int f\, d\mu$. Define the metric

$$d(\mu, \nu) = \sum_{i=1}^{\infty} \left(\mu(f_i) - \nu(f_i)\right) / (2^i \cdot ||f_i||_\infty)$$

This metric generates the weak*-topology on the space of measures $\mathcal{M}$, i.e., μ_n converges to μ in the weak star topology if and only if $d(\mu_n, \mu) \to 0$ as $n \to +\infty$.

To show compactness, it is enough to show that every infinite sequence of measures $\{\mu_n\}$ has a convergent subsequence. For each $i \geq 1$, we may select a subsequence $\mu_{n,k}$ of measures such that $\mu_{n,k}(f_j)$ converges for $1 \leq j \leq i$. The "diagonal", $\mu_{i,i}$, of this doubly indexed sequence is the desired convergent sequence. Define the linear functional $J(f_j) := \lim \mu_{i,i}(f_j)$ this is uniformly continuous on a dense subset of $C^0(X, \mathbb{R})$ and, so, has a unique extension to a bounded positive functional. Thus J arises from a measure by the Riesz Representation Theorem, which is the desired measure. □

12 Proof of Radon-Nikodym Theorem

Let us denote $\mu = m_1 + m_2$. Given $h \in L^2(X, \mu)$ we can use the Cauchy-Schwarz inequality to bound

$$|\int h dm_1| \leq \int |h| d\mu \leq \sqrt{\int |h|^2 d\mu} \sqrt{\mu(X)}$$

Thus, by the simpler and better known Riesz representation theorem for the Hilbert space $L^2(X, \mu)$, there exists $g \in L^2(X, \mu)$ such that $\int h dm_1 = \int h g d\mu$. For any $B \in \mathcal{B}$ with $\mu(B) > 0$:

$$0 < \frac{\int g \chi_B d\mu}{\mu(B)} = \frac{\mu_1(B)}{\mu(B)} \leq 1,$$

which allows us to assume that $0 \leq g(x) \leq 1$. Let $A = \{x \in X : 0 \leq g(x) < 1\}$ and for $n \geq 1$, $B \in \mathcal{B}$ let $h(x) = \chi_{A \cap B}(x)(1 + g(x) + \cdots [g(x)]^n)$. Since $\int h(1-g)dm_1 = \int hgdm_2$, we have that

$$\begin{aligned} &\int \chi_{A\cap B}(x)(1-[g(x)]^n)dm_1 \\ &= \int \chi_{A\cap B}(x)(1+g(x)+\cdots[g(x)]^{n-1})(1-g(x))dm_1(x) \\ &= \int \chi_{A\cap B}(x)(g(x)+[g(x)]^2+\cdots[g(x)]^n)dm_2(x) \end{aligned}$$

Letting $n \to +\infty$, this converges by the dominated convergence theorem to

$$m_1(A) = \int_A \left[\frac{g(x)}{1-g(x)}\chi_B(x)\right] dm_2(x)$$

This completes the proof, where $\frac{dm_1}{dm_2}(x) = \frac{g(x)}{1-g(x)}\chi_B(x)$. □

13 Convergence Theorems and Fatou's Lemma

13.1 Proof of Monotone Convergence

The limit $f_n(x) = \lim_{n\to+\infty}$ is measurable since $f^{-1}(a,b] = \cup_n f_n^{-1}(a,\infty] \in \mathcal{B}$. By definition of the integral:

$$\sup_n \int f_n d\mu = \lim_n \int f_n d\mu \leq \int f d\mu \tag{13.1}$$

Let $g(x) = \sum_i a_i \chi_{F_i} \leq f(x)$ be a simple function and let $c > 0$. Consider the sets $A_n = \{x \in X : f_n(x) \geq cg(x)\} \in \mathcal{B}$, then $A_1 \subset A_2 \subset A_3 \subset \cdots$ and $X = \cup A_n$. Then for each $n \geq 1$

$$\int f_n d\mu \geq \int_{A_n} f_n d\mu \geq c \int_{A_n} g d\mu = \sum_i a_i \mu(F_i)$$

Since $\mu(E_n \cap F_i) \to \mu(E \cap F_i)$, as $n \to +\infty$, the last term tends to $c\int gd\mu$. In particular, $\sup_n \int f_n d\mu \geq c \int gd\mu$ for $0 < c < 1$. Taking the supremum over c and then g gives that

$$\lim_{n\to+\infty} \int f_n d\mu \geq \sup \int gd\mu = \int fd\mu \tag{13.2}$$

Comparing (13.1) and (13.2) completes the proof. □

As a corollary, we get Fatou's Lemma

13.2 Proof of Fatou's Lemma

Let $F_k(x) = \inf_{n\geq k} f_n(x)$. In particular, this is an increasing sequence of functions with $F_k(x) \to \liminf_n f_n(x)$ and $F_k(x) \leq f_n(x)$, for $n \geq k$. By the monotone convergence theorem,

$$\int \liminf_n f_n(x) d\mu(x) = \lim_{k\to\infty} \int F_k(x) d\mu \leq \liminf_{n\to+\infty} \int f_n d\mu,$$

as required. □.

13.3 Proof of dominated convergence theorem

If $|f_n(x)| \leq g(x)$ then let us introduce $h_k(x) = g(x) - f_k(x) \geq 0$. In particular, $\liminf_k h_k = g - f$, by hypothesis. By the Monotone Convergence Theorem and Fatou's Lemma:

$$\int (g-f)d\mu = \int \liminf_k h_k d\mu \leq \liminf_k \int (g - f_k)d\mu = \int g d\mu - \limsup_k \int f_k d\mu. \tag{13.3}$$

Using instead $g + f_k$ in place of $g - f_k$ we have that

$$\int (g+f)d\mu \leq \int g d\mu + \liminf_k \int f_k d\mu \tag{13.4}$$

Comparing (13.3) and (13.4) we see that

$$\limsup_k \int f_k d\mu \leq \int f d\mu \leq \liminf_k \int f_k d\mu,$$

showing that $\lim_k \int f_k d\mu = \int f d\mu$.

14 Proofs of Ergorof's and Lusin's Theorems

14.1 Proof of Egoroff's Theorem

Let

$$S(n,k) = \cap_{i,j>n} \left\{ x \in X \, : \, |f_i(x) - f_j(x)| < \frac{1}{k} \right\}$$

then $S(n,k) \subset S(n+1,k)$. Since $f_m(x) \to f(x)$ as $m \to +\infty$, it is easy to see that for any $k \geq 1$ we can write $X = \cup_{n=1}^{\infty} S(n,k)$. In particular, for each $k \geq 1$, we have $\mu(S(n,k)) \to 1$, as $n \to +\infty$, and so we can choose n_k sufficiently large that $\mu(S(n_k,k)) > 1 - \frac{\epsilon}{2^k}$.

We define $Y = \cap_n S(n_k, k)$ and then

$$\mu(X - Y) \leq \sum_k \mu(X - S(n_k,k)) \leq \sum_k \frac{\epsilon}{2^k} = \epsilon.$$

It remains to show that $f_n \to f$ uniformly on Y. Given $\delta > 0$, choose $1/k < \delta$. If $x \in E \subset S(n, n_k)$ then by definition $|f_i(x) - f_j(x)| < \delta$ for all $i, j > n_k$. This gives the required uniform convergence.

Let us begin by recalling a well known result in topology:

Lemma 14.1 (Urysohn's Lemma)**.** *Let X be a compact metric space and $K \subset U \subset X$, where K is closed and U is open. Then there exists a continuous function $f : X \to \mathbb{R}$ such that*

$$h(x) = \begin{cases} 1 & \text{if } x \in K \\ 0 & \text{if } x \in X - U. \end{cases}$$

This is useful in the proof of the following.

14.2 Proof of Lusin's Theorem

Assume first that $0 \leq f < 1$. We can choose an increasing sequence of (positive) simple functions s_n such that

1. $s_n(x) \to f(x)$;
2. $t_n(x) := s_n(x) - s_{n-1}(x) = 0$ or 2^{-n}.

Define $t_1 = s_1$, $t_n = s_n - s_n - 1$, $n \geq 2$ then $f(x) = \sum_{n=1}^{\infty} t_n(x)$. Let $T_n = \{x : t_n(x) = 2^{-n}\}$. There are compact subsets K_n and open subsets V_n such that $K_n \subset T_n \subset V_n \subset X$ such that $m(V_n - K_n) < 2^{-n}\epsilon$, for all $n \geq 1$. By Urysohn's lemma there is a continuous functions $h_n : X \to [0, 1]$ such that

$$h_n(x) = \begin{cases} 1 \text{ on } K_n \text{ and} \\ 0 \text{ for } x \notin \overline{V_n} \end{cases}$$

Define $g(x) = \sum_n 2^{-n} h_n(x)$, for $x \in X$. By uniform convergence, $g(x)$ is continuous, and $g(x) = 0$ for $x \in X - \overline{V}$. Furthermore, $2^{-n} h_n(x) = t_n(x)$, except for $x \in V_n - K_n$, so we see that $g(x) = f(x)$ except for $x \in \cup_n (V_n - K_n)$ and this union has measure at most ϵ. We then let $Y = X - \cup_n (V_n - K_n)$.

If f is essentially bounded then this argument is easily modified. More generally, we can let $f_n(x) = \min\{n, f(x)\}$, for $n \geq 1$. If we choose n sufficiently large then f_n and f agree on a set of measure at most ϵ. We can then apply the above argument to f_n. □