Measure Theory, 2010

June 25, 2010

Contents

1 Set theoretical background

As with much of modern mathematics, we will be using the language of set theory and the notation of logic.

Let X be a set.

We say that Y is a *subset* of X if every element of Y is an element of X. I will variously denote this by $Y \subset X$, $Y \subseteq X$, $X \supset Y$, and $X \supseteq Y$. In the case that I want to indicate that Y is a *proper* subset of X – that is to say, Y is a subset of X but it is not equal to X – I will use $Y \subsetneq X$.

We write $x \in X$ or $X \ni x$ to indicate that x is an element of the set X. We write $x \notin X$ to indicate that x is *not* an element of X.

For $P(\cdot)$ some property we use $\{x \in X : P(x)\}$ or $\{x \in X | P(x)\}$ to indicate the set of elements $x \in X$ for which $P(x)$ holds. Thus for $x \in X$, we have $x \in \{x \in X : P(x)\}$ if and only if $P(x)$.

There are certain operations on sets. Given $A, B \subset X$ we let $A \cap B$ be the intersection – and thus $x \in A \cap B$ if and only if x is an element of A *and* x is an element of B. We use $A \cup B$ for the union – so $x \in A \cup B$ if and only if x is a member of at least one of the two. We use $A \setminus B$ to indicate the elements of A which are *not* elements of B. In the case that it is understood by context that all the sets we are currently considering are subsets of some fixed set X, we use A^c to indicate the complement of A in X – in other words, $X \setminus A$. $A \Delta B$ denotes the *symmetric difference* of A and B, the set of points which are in one set but not the other. Thus $A \Delta B = (A \setminus B) \cup (B \setminus A)$. We use $A \times B$ to indicate the set of all pairs (a, b) with $a \in A$, $b \in B$. $\mathcal{P}(X)$, the *power set of* X, indicates the set of *all* subsets of X – thus $Y \in \mathcal{P}(X)$ if and only if $Y \subset X$. We let B^A, also written

$$\prod_A B,$$

be the collection of all functions from A to B.

A very, very special set is the empty set: $\emptyset$. It is the set which has no members. If you like, it is the characteristically zen set. Some other special sets are: $\mathbb{N} = \{1, 2, 3, ...\}$, the set of natural numbers; $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$, the set of integers; $\mathbb{Q}$, the set of rational numbers; $\mathbb{R}$ the set of real numbers.

Given some collection $\{Y_\alpha : \alpha \in \Lambda\}$ of sets, we write

$$\bigcup_{\alpha \in \Lambda} Y_\alpha$$

or

$$\bigcup \{Y_\alpha : \alpha \in \Lambda\}$$

to indicate the union of the Y_α's. Thus $x \in \bigcup_{\alpha \in \Lambda} Y_\alpha$ if and only if there is some $\alpha \in \Lambda$ with $x \in Y_\alpha$. A slight variation is when we have some property $P(\cdot)$ which could apply to the elements of Λ and we write

$$\bigcup \{Y_\alpha : \alpha \in \Lambda, P(\alpha)\}$$

or

$$\bigcup_{\alpha \in \Lambda, P(\alpha)} Y_\alpha$$

to indicate the union over all the Y_α's for which $P(\alpha)$ holds. The obvious variations hold on this for intersections. Thus we use

$$\bigcap_{\alpha \in \Lambda} Y_\alpha$$

or

$$\bigcap \{Y_\alpha : \alpha \in \Lambda\}$$

to indicate the set of x which are members *every* Y_α. Given a infinite list $(Y_\alpha)_{\alpha\in\Lambda}$ of sets, we use

$$\prod_{\alpha\in\Lambda} Y_\alpha$$

to indicate the infinite product – which formerly may be thought of as the collection of all functions $f : \Lambda \to \bigcup_{\alpha\in\Lambda} Y_\alpha$ with $f(\alpha) \in Y_\alpha$ at every α.

Given two sets X, Y and a function

$$f : X \to Y$$

between the sets, there are various set theoretical operations involved with the function f. Given $A \subset X$, $f|_A$ indicates the function from A to Y which arises from the restriction of f to the smaller domain. Given $B \subset Y$ we use

$$f^{-1}[B]$$

to indicate the *pullback* of B along f – in other words, $\{x \in X : f(x) \in B\}$. Given $A \subset X$ we use $f[A]$ to indicate the *image* of A – the set of $y \in Y$ for which there exists some $x \in A$ with $f(x) = y$. Given $A \subset X$ we use χ_A to denote the *characteristic function* or *indicator function* of A. This is the function from X to $\mathbb{R}$ which assumes the value 1 at each element of A and the value 0 on each element of A^c.

A function $f : X \to Y$ is said to be *injective* or *one-to-one* if whenever $x_1, x_2 \in X$ with $x_1 \neq x_2$ we have $f(x_1) \neq f(x_2)$ – different elements of X move to different elements of Y. The function is said to be *surjective* or *onto* if every element of Y is the image of some point under f – in other words, for any $y \in Y$ we can find some $x \in X$ with $f(x) = y$. The function $f : X \to Y$ is said to be a *bijection* or a *one-to-one correspondence* if it is both an injection and a surjection. In this case of a bijection, we can define $f^{-1} : Y \to X$ by the requirement that $f^{-1}(y) = x$ if and only if $f(x) = y$.

We say that two sets *have the same cardinality* if there is a bijection between them. Note here that the inverse of a bijection is a bijection, and thus if A has the same cardinality as B (i.e. there exists a bijection $f : A \to B$), then B has the same cardinality as A. The composition of two bijections is a bijection, and thus if A has the same cardinality as B and B has the same cardinality as C, then A has the same cardinality as C.

In the case of finite sets, the definition of cardinality in terms of bijections accords with our commonsense intuitions – for instance, if I count the elements in set of "days of the week", then I am in effect placing that set in to a one to one correspondence with the set $\{1, 2, 3, 4, 5, 6, 7\}$. This theory of cardinality can be extended to the realm of the infinite with unexpected consequences.

A set is said to be *countable* if it is either finite or it can be placed in a bijection with $\mathbb{N}$, the set of natural numbers.[1]. A set is said to have *cardinality* $\aleph_0$ if it can be placed in a bijection with $\mathbb{N}$. Typically we write $|A|$ to indicate the cardinality of the set A.

Lemma 1.1 *If $A \subset \mathbb{N}$, then A is countable.*

Proof If A has no largest element, then we define $f : \mathbb{N} \to A$ by $f(n) = n^{\text{th}}$ element of A. □

Corollary 1.2 *If A is a set and $f : A \to \mathbb{N}$ is an injection, then A is countable.*

Lemma 1.3 *If A is a set and $f : \mathbb{N} \to A$ is a surjection, then A is countable.*

Proof Let B be the set

$$\{n \in \mathbb{N} : \forall m < n(f(n) \neq f(m))\}.$$

B is countable, by the last lemma, and A admits a bijection with B. □

[1] But be warned: A small minority of authors only use *countable* to indicate a set which can be placed in a bijection with $\mathbb{N}$

Corollary 1.4 *If A is countable and $f : A \to B$ is a surjection, then B is countable.*

Lemma 1.5 $\mathbb{N} \times \mathbb{N}$ *is countable.*

Proof Define

$$f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$$
$$(m, n) \mapsto 2^n 3^m.$$

This is injective, and then the result follows the corollary to 1.1. □

Lemma 1.6 *The countable union of countable sets is countable.*

Proof Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of countable sets; the case that we have a finite sequence of countable sets is similar,but easier. We may assume each A_n is non-empty, and then at each n fix a surjection $\pi_n : \mathbb{N} \to \mathbb{N}$. Then

$$f : \mathbb{N} \times \mathbb{N} \to \bigcup_{n\in\mathbb{N}} A_n$$
$$(n, m) \mapsto \pi_n(m)$$

gives a surjection from $\mathbb{N} \times \mathbb{N}$ onto the union. Then the lemma follows from 1.3 and 1.5. □

Lemma 1.7 $\mathbb{Z}$ *is countable.*

Proof Since there is a surjection of two disjoint copies of $\mathbb{N}$ onto $\mathbb{Z}$, this follows from 1.6. □

Lemma 1.8 $\mathbb{Q}$ *is countable.*

Proof Let $A = \mathbb{Z} \setminus \{0\}$. Define

$$\pi : \mathbb{Z} \times A \to \mathbb{Q}$$
$$(\ell, m) \mapsto \frac{\ell}{m}.$$

π is a surjection, and so the lemma follows from 1.3 □

Seeing this for the first time, one might be tempted to assume that all sets are countable. Remarkably, no.

Theorem 1.9 $\mathbb{R}$ *is uncountable.*

Proof For a real number $x \in [0, 1]$, let $f_n(x)$ be the n^{th} digit in its decimal expansion. Note then that for any sequence $(a_n)_{n\in\mathbb{N}}$ with each

$$a_n \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\},$$

we can find a real number in the unit interval $[0, 1]$ with $f_n(x) = a_n$ – except in the somewhat rare case that the a_n's eventually are all equal to 9.

Now let $h : \mathbb{N} \to \mathbb{R}$ be a function. It suffices to show it is not surjective. And for that purpose it suffices to find some x such that at every n there exists an m with $f_m(x) \neq f_m(h(n))$ – in other words to find an x whose decimal expansion differs from each $h(n)$ at some m.

Now define $(a_n)_n$ by the requirement that $a_n = 5$ if $f_n(h(n)) \geq 6$ and $a_n = 6$ if $f_n(h(n)) \leq 5$. For x with the a_n's as its decimal expansion, in other words

$$f_n(x) = a_n$$

at every n, we have

$$\forall n \in \mathbb{N}(f_n(h(n)) \neq f_n(x)),$$

and we are done. □

Exercise The proof given above of 1.3 implicitly used prime factorization – to the effect that

$$2^n 3^m = 2^i 3^j$$

implies $n = i$ and $m = j$. Try to provide a proof which does not use this theorem.

Exercise Show $|\mathbb{R} \times \mathbb{R}| = |\mathbb{R}|$.

Exercise Show that $\mathbb{N}^{<\infty}$, the collection of all finite sequences from $\mathbb{N}$, is countable.

Exercise Show that $\mathcal{P}(\mathbb{N})$ and $\mathbb{N}^{\mathbb{N}}$ are uncountable. (Hint: Much the same as the proof that $\mathbb{R}$ is uncountable.)

Exercise Show that $\mathcal{P}(\mathbb{N})$ and $\mathbb{N}^{\mathbb{N}}$ have the same cardinality as $\mathbb{R}$. (Harder).

Theorem 1.10 *If A is a set, then $|A| < |\mathcal{P}(A)|$.*

Proof First note that

$$x \mapsto \{x\}$$

gives an injection from A to $\mathcal{P}(A)$.

Conversely, suppose $f : A \to \mathcal{P}(A)$ is a function. We wish to show it is not surjective. So define

$$B = \{x : x \notin f(x)\}.$$

Now suppose $B = f(x)$ some $x \in A$. Then

$$x \in f(x) \Leftrightarrow x \in B \Leftrightarrow x \notin f(x).$$

with a contradiction. □

A somewhat more complete introduction to the theory of infinite cardinals can be found in [12].

At some point also we will need to avail ourselves of "Zorn's lemma".

Theorem 1.11 *(Zorn's lemma) Let X be a set equipped with a partial order $\leq$. Assume that whenever $A \subset X$ is linearly ordered by $\leq$ then it has an upper bound – i.e.*

$$\exists x \in X \forall a \in A (a \leq x).$$

Then X has a maximal element – i.e.

$$\exists x \in X \forall y \in X (x \leq y \Rightarrow x = y).$$

Here a set $(X, \leq)$ is said to be a *partially ordered set* if: $\forall a \in X (a \leq a)$ (reflexivity); $\forall a, b \in X ((a \leq b \wedge b \leq a) \Rightarrow a = b)$ (antisymmetry); $\forall a, b, c \in X ((a \leq b \wedge b \leq c) \Rightarrow a \leq c)$ (transitivity). If in addition we have $\forall a, b \in X (a \leq b \vee b \leq a)$ then we say that $(X, \leq)$ is *linearly ordered*.

2 Review of topology, metric spaces, and compactness

On the whole these notes presuppose a first course in metric spaces. This is only a quick review, with a special emphasis on the aspects of compactness which will be especially relevant when we come to consider $C(K)$ in the chapter on the Reisz representation theorem. A thorough, more complete, and far better introduction to the subject can be found in [4] or [9]. There is a substantial degree of abstraction in first passing from the general properties of distance in euclidean spaces to the notion of a general metric space, and then the notion of a general topological space; this short chapter is hardly a sufficient guide.

Definition A set X equipped with a function

$$d : X \times X \to \mathbb{R}$$

is said to be a *metric space* if

1. $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;
2. $d(x, y) = d(y, x)$;
3. $d(x, y) \leq d(x, z) + d(z, y)$.

The classic example of a metric is in fact the reals, with the euclidean metric

$$d(x, y) = |x - y|.$$

Then $d(x, y) \leq d(x, z) + d(z, y)$ is the triangle inequality. A somewhat more exotic example would be take any random set X and let $d(x, y) = 1$ if $x \neq y$, $= 0$ otherwise.

Definition A sequence $(x_n)_{n\in\mathbb{N}}$ of points in a metric space (X, d) is said to be *Cauchy* if

$$\forall \epsilon > 0 \exists N \forall n, m > N(d(x_n, x_m) < \epsilon).$$

A sequence is said to*converge* to a limit $x_\infty \in X$ if

$$\forall \epsilon > 0 \exists N \forall n > N(d(x_n, x_\infty) < \epsilon).$$

This is often written as

$$x_n \to x_\infty,$$

and we then say that x_∞ is the *limit* of the sequence. We say that $(x_n)_n$ is *convergent* if it converges to some point.

A metric space is said to be *complete* if every Cauchy sequence is convergent.

Lemma 2.1 *A convergent sequence is Cauchy.*

Theorem 2.2 *Every metric space can be realized as a subspace of a complete metric space.*

For instance, $\mathbb{Q}$ with the usual euclidean metric is *not* complete, but it sits inside $\mathbb{R}$ which is.

Definition If (X, d) is a metric space, then for $\epsilon > 0$ and $x \in X$ we let

$$B_\epsilon(x) = \{y \in X : d(x, y) < \epsilon\}.$$

We then say that $V \subset X$ is *open* if for all $x \in V$ there exists $\epsilon > 0$ with

$$B_\epsilon(x) \subset V.$$

A subset of X is *closed* if its complement is open.

Lemma 2.3 *A subset A of a metric space X is closed if and only if whenever $(x_n)_n$ is a convergent sequence of points in A the limit is in A.*

Definition For (X, d) and (Y, ρ) two metric spaces, we say that a function

$$f : X \to Y$$

is *continuous* if for any open $W \subset Y$ we have $f^{-1}[W]$ open in X.

The connection between this definition and the customary notion of continuous function for $\mathbb{R}$ is made by the following lemma:

Lemma 2.4 *A function $f : X \to Y$ between the metric space (x, d) and the metric space (Y, ρ) is continuous if and only if for all $x \in X$ and $\delta > 0$ there exists $\epsilon > 0$ such that*

$$\forall x' \in X(d(x, x') < \epsilon \Rightarrow \rho(f(x), f(x')) < \delta).$$

Definition For (X, d) and (Y, ρ) two metric spaces, we say that a function

$$f : X \to Y$$

is *uniformly continuous* if for all $\delta > 0$ there exists $\epsilon > 0$ such that

$$\forall x, x' \in X(d(x, x') < \epsilon \Rightarrow \rho(f(x), f(x')) < \delta)).$$

In the definitions above we have various notions which are built around the concept of *open set*. It turns out that this key idea admits a powerful generalization and abstraction – we can talk about "spaces" where there is a notion of open set, but no notion of a metric.

Definition A set X equipped with a collection $\tau \subset \mathcal{P}(X)$ is said to be a *topological space* if:

1. $X, \emptyset \in \tau$;
2. τ is closed under finite intersections – $U, V \in \tau \Rightarrow U \cap V \in \tau$;
3. τ is closed under arbitrary unions – $S \subset \tau \Rightarrow \bigcup S \in \tau$.

In this situation, we say call the elements of τ *open sets.*

Lemma 2.5 *If (X, d) is a metric space, then the sets which are open in X (in our previous sense) form a topology on X.*

Lemma 2.6 *Let X be a set and $B \subset \mathcal{P}(X)$ which is closed under finite unions and includes the empty set. Suppose additionally that $\bigcup B = X$. Then the collection of all arbitrary unions from X forms a topology on X.*

Definition If $B \subset \mathcal{P}(X)$ is as above, and τ is the resulting topology, then we say that B is a *basis for τ*.

Lemma 2.7 *Let $(X_i, \tau_i)_{i \in I}$ be an indexed collection of topological spaces. Then there is a topology on*

$$\prod_{i \in I} X_i$$

with basic open sets of the form

$$\{f \in \prod_{i \in I} X_i : f(i_1) \in V_1, f(i_2) \in V_2, ...f(i_N) \in V_N\},$$

where $N \in \mathbb{N}$ and each of $V_1, V_2, ...V_N$ are open in the respective topological spaces $X_{i_1}, X_{i_2}, ..., X_{i_N}$.

Definition In the situation of the above lemma, the resulting topology is called *the product topology.*

Definition Let X be a topological space and $C \subset X$ a subset. Then the *subspace topology* on C is the one consisting of all subsets of the form $C \cap V$ where V is open in X.

Technically one should prove before this definition that the subspace topology *is* a topology, but that is trivial to verify. Quite often people will simply view a subset of a topological space as a topological space in its own right, without explicitly specifying that they have in mind the subspace topology.

Definition In a topological space X we say that $A \subset X$ is *compact* if whenever

$$\{U_i : i \in I\}$$

is a collection of open sets with

$$A \subset \bigcup_{i \in I} U_i$$

we have some finite $F \subset I$ with

$$A \subset \bigcup_{i \in F} U_i.$$

In other words, every *open cover* of A has a finite subcover. X is said to be a *compact space* if we obtain that X is compact as a subset of itself – namely, every open cover of X has a finite subcover.

Examples

1. Any finite subset of a metric space is compact. Indeed, one should think of compactness as a kind of topological generalization of finiteness.
2. $(0, 1)$, the open unit interval, is *not* compact in $\mathbb{R}$ under the usual euclidean metric, even though it does admit *some finite* open covers, such as $\{(0, \frac{1}{2}), (\frac{1}{4}, 1)\}$. Instead if we let $U_n = (\frac{1}{n+2}, \frac{1}{n})$, then $\{U_n : n \in \mathbb{N}\}$ is an open cover without any finite subcover.
3. Let X be a metric space which is not *complete*. Then it is not compact. (Let $(x_n)_n$ be a Cauchy sequence which does not converge in X. Let Y be a larger metric space including X to which the sequence converges to some point x_∞. Then at each n let U_n be the set of points *in* X which have distance greater than $\frac{1}{n}$ from x_∞.)

Definition Let X and Y be topological spaces. A function

$$f : X \to Y$$

is said to be *continuous* if for any set $U \subset Y$ which is open in Y we have

$$f^{-1}[V]$$

open in X.

Lemma 2.8 *Let X be a compact space and $f : X \to \mathbb{R}$ continuous. Then f is bounded.*

In fact there is much more that can be said:

Theorem 2.9 *The following are equivalent for a metric space X:*

1. *It is compact.*

2. *It is complete and totally bounded – i.e. for every $\epsilon > 0$ we can cover X with finitely many balls of the form $B_\epsilon(x)$.*

3. *Every sequence has a convergent subsequence – i.e. X is "sequentially compact".*

4. *Every continuous function from X to $\mathbb{R}$ is bounded.*

5. *Every continuous function from X to $\mathbb{R}$ is bounded and attains its maximum.*

Some of this beyond the scope of this brief review, and I will take the equivalence of 2 and 3 as read. However, it might be worth briefly looking at the equivalence of 2 and 3 with 4. I will start with the implication from 4 to 2.

First suppose that there is a Cauchy sequence $(x_n)_n$ which does not converge. Appealing to 2.2, let Y be a larger metric space in which $x_n \to x_\infty$. Then define

$$f : X \to \mathbb{R}$$

by

$$x \mapsto \frac{1}{d(x, x_\infty)}.$$

The function is well defined on X since every point in X will have some positive distance from x_∞. It is a minor exercise in $\epsilon - \delta$-ology to verify that the function is continuous, but in rough terms it is because if x, x' are two points which are in X and sufficiently close to each other, relative to their distance to x_∞, then in Y the values

$$\frac{1}{d(x, x_\infty)}, \frac{1}{d(x', x_\infty)}$$

will be close.

Now suppose that the metric space is not totally bounded. We obtain some $\epsilon > 0$ such that no finite number of ϵ balls covers X. From this we can get that the are infinitely many disjoint $\epsilon/2$ balls,

$$B_{\epsilon/2}(z_1), B_{\epsilon/2}(z_2), ...$$

Then we let U be the union of these open balls. We define f to be 0 on the complement of U. Inside each ball we define f separately, with

$$f(x) = n \cdot d(x, X \setminus U)$$
$$=_{\text{df}} n \cdot (\inf_{z \in X \setminus U} d(x, z)).$$

The function takes ever higher peaks inside the $B_{\epsilon/2}(x_n)$ balls, and thus has no bound. The balls in which the function is non-zero are sufficiently spread out in the space that we only need to verify that f is continuous on each $B_{\epsilon/2}(z_i)$, which is in turn routine.

For 3 implying 4, suppose $f : X \to \mathbb{R}$ has no bound. Then at each n we can find x_n with $f(x_n) > n$. Going to a convergent subsequence we would be able to assume that $x_n \to x_\infty$ for some $x_\infty \in X$, but then there would be no value for $f(x_\infty)$ which would allow the function to remain continuous.

It actually takes some serious theorems to show that there are *any* compact spaces.

Theorem 2.10 *(Tychonov's theorem) Let $(X_i, \tau_i)_{i \in I}$ be an indexed collection of compact topological spaces. Then*

$$\prod_{i \in I} X_i$$

is a compact space in the product topology.

Proof Let us say that an open $V \subset \prod_{i \in I} X_i$ is *subbasic* if it has the form

$$\{f : f(i) \in U_i\}$$

for some single $i \in I$ and open $U \subset X_i$. Note then that every basic open set is a finite intersection of subbasic open sets.

Claim: Let

$$\mathcal{S} \cup \{V_1 \cap V_2 \cap ... \cap V_n\}$$

be a collection of open sets for which some no finite subset covers X. Assume each V_i is subbasic. Then for some $i \leq n$ no finite subset of

$$\mathcal{S} \cup \{V_i\}$$

covers X.

Proof of Claim: Suppose at each i we have some finite $\mathcal{S}_i \subset \mathcal{S}$ such that

$$\mathcal{S}_i \cup \{V_i\}$$

covers X. Then

$$\mathcal{S}_1 \cup \mathcal{S}_2 \cup ... \cup \mathcal{S}_n \cup \{V_1 \cap V_2 \cap ... V_n\}$$

covers X. (□Claim)

So now let $\mathcal{S}$ be a collection of open sets such that no finite subset covers. We may assume $\mathcal{S}$ consists only of basic open sets. Then applying the above claim we can steadily turn each basic open set into a subbasic open set.[2] so that at last we obtain some $\mathcal{S}^*$ with

1. $$\bigcup \mathcal{S} \subset \bigcup \mathcal{S}^*;$$

2. no finite subset of $\mathcal{S}^*$ of covers X;

3. $\mathcal{S}^*$ consists solely of subbasic open sets.

Then at each i, we can appeal to the compactness of X_i and obtain some $x_i \in X_i$ such that for every open $V \subset X$ with

$$V \times \prod_{j \in I, j \neq i} X_j$$

we have $x_i \notin V_i$. But if we let $f \in \prod_{i \in I} X_i$ be defined by

$$f(i) = x_i$$

then we obtain an element of the product space not in the union of $\mathcal{S}^*$, and hence not in the union of $\mathcal{S}$. □

Theorem 2.11 *(Heine-Borel) The closed unit interval*

$$[0, 1]$$

is compact. More generally, any subset of $\mathbb{R}$ *is compact if and only if it is closed and bounded.*

[2] In fact one needs a specific consequence of the axiom of choice called *Zorn's lemma* to formalize this part of the proof precisely

In particular, if $a < b$ are in $\mathbb{R}$ and we have a sequence of open intervals of the form (c_n, d_n) with

$$[a,b] \subset \bigcup_{n\in\mathbb{N}} (c_n, d_n),$$

then there is some finite N with

$$[a,b] \subset \bigcup_{n\leq N} (c_n, d_n).$$

Here we say that a subset $A \subset \mathbb{R}$ is bounded if there is some positive c with

$$|x| < c$$

all $x \in A$.

Lemma 2.12 *The continuous image of a compact space is compact.*

Proof If X is compact, $f : X \to Y$ continuous, let $C = f[X]$. Then for any open covering $\{U_i : i \in I\}$ of C, we can let

$$V_i = f^{-1}[U_i]$$

at each $i \in I$. From the definition of f being continuous, each V_i is open. Since the U_i's cover $C = f[X]$, the V_i's cover X. Since X is compact, there is some finite $F \subset I$ with

$$X \subset \bigcup_{i\in F} V_i,$$

which entails

$$C \subset \bigcup_{i\in F} U_i.$$

□

Lemma 2.13 *A closed subset of a compact space is compact.*

For us, the main consequences of compactness are for certain classes of function spaces. The primary examples will be Ascoli-Arzela and Alaoglu. Both these theorems are fundamentally appeals to Tychonov's theorem, and can be viewed as variants of the following observations:

Definition Let (X, d) and (Y, ρ) be metric spaces. Let $C(X, Y)$ be the space of continuous functions from X to Y. Define

$$D : C(X,Y)^2 \to \mathbb{R} \cup \{\infty\}$$
$$D(f,g) = \sup_{z\in X} \rho(f(z), g(z)).$$

This metric D, in the cases *when it is a metric*, is sometimes called the *sup norm metric.*

Lemma 2.14 *Let (X, d) and (Y, ρ) be metric spaces. If either X or Y is compact, then D always takes finite values and forms a metric.*

Proof The characteristic properties of being a metric are clear, once we show D is finite. The finiteness of D is clear in the case that Y is compact, since the metric on Y will be totally bounded, and hence there will be a single $c > 0$ such that for all $y, y' \in Y$ we have $\rho(y, y') < c$.

On the other hand, suppose X is compact. Then for any two continuous functions $f, g : X \to Y$ we have $C = f[X] \cup g[X]$ compact by 2.12. But then since this set is ϵ-bounded for all $\epsilon > 0$ we in particular have

$$\sup_{y,y'\in C} \rho(y, y') < \infty$$
$$\therefore \sup_{z\in X} \rho(f(z), g(z)) < \infty.$$

□

Lemma 2.15 *Let (X, d) and (Y, ρ) be metric spaces. Assume X is compact. If $f : X \to Y$ is continuous, then it is uniformly continuous.*

Proof Given $\epsilon > 0$, we can find at each $x \in X$ some $\delta(x)$ such that any two points in $B_{\delta(x)}(x)$ have images within ϵ under f. Then at each x let $V_x = B_{\delta(x)/2}(x)$. The V_x's cover X, and then by compactness there is finite $F \subset X$ with

$$X = \bigcup_{x \in F} V_x.$$

Taking $\delta = \min\{\delta(x)/2 : x \in F\}$ completes the proof. □

Definition In the above, let $C_1(X, Y)$ be the elements f of $C(X, Y)$ with

$$d(x, y) \geq \rho(f(x), f(y))$$

all $x, y \in X$.

Theorem 2.16 *If (X, d) and (Y, ρ) are both compact metric spaces, then $C_1(X, Y)$ is compact.*

Proof The first and initially surprising fact to note is that the sup norm induces the same topology on $C_1(X, Y)$ as the topology it induces as a subspace of

$$\prod_X Y$$

in the product topology. The point here is that $F \subset X$ is an ϵ-net – which is to say, and point in X is within distance ϵ of some element of F – and $f, g \in C_(X, Y)$ have $\rho(f(x), g(x)) < \epsilon$ for all $x \in F$, then $D(f, g) \leq 3\epsilon$.

Now the proof is completed by Tychonov's theorem once we observe that $C_1(X, Y)$ is a closed subspace of $\prod_X Y$ in the product topology. □

It is only a slight exaggeration to say that Ascoli-Arzela and Alaoglu are corollaries of 2.16. The statements of the two theorems are more specialized, but the proofs are almost identical.

Finally for the work on the Riesz representation theorem, it is important to know that in some cases $C(X, Y)$ will form a *complete* metric space.

Theorem 2.17 *Let (X, d) and (Y, ρ) be metric spaces. Suppose that X is compact and Y is complete. Then $C(X, Y)$ is a complete metric space.*

Proof Suppose $(f_n)_{n \in \mathbb{N}}$ is a sequence which is Cauchy with respect to D. Then in particular at $x \in X$ the sequence

$$(f_n(x))_{n \in X}$$

is Cauchy in Y, and hence converges to some value we will call $f(x)$. It remains to see that

$$f : X \to Y$$

is continuous.

Fix $\epsilon > 0$. If we go to N with

$$\forall n, m \geq N(D(f_n, f_m) < \epsilon)$$

then $\rho(f_N(x), f(x)) \leq \epsilon$ all $x \in X$. Then given a specific $x \in X$ we can find $\delta > 0$ such that for any $x' \in B_\delta(x)$ we have $\rho(f_N(x), f_N(x')) < \epsilon$, which in turn implies $\rho(f(x), f(x')) < 3\epsilon$. □

3 The concept of a measure

Definition For S a set we let $\mathcal{P}(S)$ be the set of all subsets of S. $\Sigma \subset \mathcal{P}(S)$ is an *algebra* if it is closed under complements, unions, and intersections; it is a *σ-algebra* if it is closed under complements, *countable* unions, and *countable* intersections.

Here by a countable union we mean one of the form $\bigcup_{n\in\mathbb{N}} A_n$ and by countable intersection one of the form $\bigcap_{n\in\mathbb{N}} A_n$.

Definition A set S equipped with a σ-algebra Σ is said to be a *measure space* if $\Sigma \subset \mathcal{P}(S)$ is a (non-empty) σ-algebra. A function

$$\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$$

is a *measure* if

1. $\mu(\emptyset) = 0$, and
2. μ is *countably additive* – given a sequence $(A_n)_{n\in\mathbb{N}}$ of disjoint sets in Σ we have

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

Here $\mathbb{R}^{\geq 0}$ refers to the non-negative reals. We require – at least at this stage – that our measures return non-negative numbers, with the possible inclusion of positive infinity. In this definition, some sets may have infinite measure.

Exercise (a) Show that if Σ is a non-empty σ-algebra on S then S and $\emptyset$ (the empty set) are both in Σ.

(b) Show that if $\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ is countably additive and *finite* (that is to say, $\mu(A) \in \mathbb{R}$ all $A \in \Sigma$), then $\mu(\emptyset) = 0$.

The very first issue which confronts us is the existence of measures. The definition is in the paragraph above – bold and confident – but without the slightest theoretical justification that there are any non-trivial examples.

Simply taking enough classes in calculus one might develop the intuition that Lebesgue measure has the required property of σ-additivity. We will prove a sequence of entirely abstract lemmas which will show that Lebesgue measure is indeed a measure in our sense. The approach in this section will be through the Carathéodory extension theorem. Much later in the notes we will see a different approach to the existence of measures in terms of the Riesz representation theorem for continuous functions on compact spaces. Although the Riesz representation theorem could be used to show the existence of Lebesgue measure, the approach is far more abstract, appealing to various ideas from Banach space theory, and the actual verifications involved in the proof take far longer.

The proof in this section may already seem rather abstract – and in some sense, it is. Still it is an easier first pass at the notions than the path through Riesz.

Definition Let S be a set. A function

$$\lambda : \mathcal{P}(S) \to \mathbb{R}^{\geq 0} \cup \{\infty\}$$

is an *outer measure* if $\lambda(\emptyset) = 0$ and whenever

$$A \subset \bigcup_{n\in\mathbb{N}} A_n$$

then

$$\lambda(A) \leq \sum_{n\in\mathbb{N}} \lambda(A_n).$$

Lemma 3.1 *Let S be a set and suppose $K \subset \mathcal{P}(S)$ is such that for every $A \subset S$ there is a countable sequence $(A_n)_{n\in\mathbb{N}}$ with:*

1. *each $A_n \in K$;*
2. *$A \subset \bigcup_{n\in\mathbb{N}} A_n$.*

Let $\rho : K \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ with $\rho(\emptyset) = 0$.

Then if we define

$$\lambda : \mathcal{P}(S) \to \mathbb{R}^{\geq 0} \cup \{\infty\}$$

by letting $\lambda(A)$ equal the infinum of the set

$$\{\sum_{n\in\mathbb{N}} \rho(A_n) : A \subset \bigcup_{n\in\mathbb{N}} A_n; \text{ each } A_n \in K\},$$

then λ is an outer measure.

Proof This is largely an unravelling of the definitions.

The issue is to check that if $\lambda(B_n) = a_n$ and $B \subset \bigcup_{n\in\mathbb{N}} B_n$, then

$$\lambda(B) \leq \sum_{n\in\mathbb{N}} \lambda(B_n) = \sum_{n\in\mathbb{N}} a_n.$$

However if we fix $\epsilon > 0$ and if at each n we fix a covering $(B_{n,m})_{m\in\mathbb{N}}$ with

$$B_n \subset \bigcup_{m\in\mathbb{N}} B_{n,m},$$

$$\sum_m \rho(B_{n,m}) < a_n + \epsilon 2^{-m-n-1},$$

then $\bigcup_{n,m\in N} B_{n,m} \supset B$ and

$$\sum_{n,m\in N} \rho B_{n,m} < \epsilon + \sum_n \rho(a_n).$$

□

Definition Given an outer measure $\lambda : \mathcal{P}(S) \to \mathbb{R}^{\geq 0} \cup \{\infty\}$, we say that $A \subset S$ is λ-*measurable* if for any $B \subset S$ we have

$$\lambda(B) = \lambda(B \cap A) + \lambda(B \setminus A).$$

Here $B \setminus A$ refers to the elements of B not in A. If we adopt the convention that A^c is the *relative complement of A in S* – the elements of S not in A – then we could as well write this as $B \cap A^c$.

Theorem 3.2 *(Carathéodory extension theorem, part I) Let λ be an outer measure on S and let Σ be the collection of all λ-measurable sets. Then Σ is a σ-algebra and λ is a measure on Σ.*

Proof The closure of Σ under complements should be clear from the definitions.

Before doing closure under countable unions and intersections, let us first do finite intersections. For that purpose, it suffices to do intersections of size two, since any finite intersection can be obtained by repeating the operation of a single intersection finitely many times.

Suppose $A_1, A_2 \in \Sigma$ and $B \subset S$. Applying our assumptions on A_1 we obtain

$$\lambda(B \cap (A_1 \cap A_2)^c) = \lambda((B \cap A_1^c) \cup (B \cap A_2^c)) =$$

$$\lambda((B \cap A_1^c) \cup (B \cap A_2^c)) \cap A_1^c) + \lambda((B \cap A_1^c) \cup (B \cap A_2^c)) \cap A_1) = \lambda(B \cap A_1^c) + \lambda(B \cap A_1 \cap A_2^c).$$

Then applying the assumptions on A_1 once more we obtain

$$\lambda(B) = \lambda(B \cap A_1) + \lambda(B \cap A_1^c),$$

and then applying assumptions on A_2, this equals

$$\lambda(B \cap A_1 \cap A_2) + \lambda(B \cap A_1 \cap A_2^c) + \lambda(B \cap A_1^c),$$

which after using $\lambda(B \cap (A_1 \cap A_2)^c) = \lambda(B \cap A_1^c) + \lambda(B \cap A_1 \cap A_2^c)$ from above gives

$$\lambda(B) = \lambda(B \cap A_1 \cap A_2) + \lambda(B \cap (A_1 \cap A_2)^c),$$

as required.

Having Σ closed under complements and finite intersections we obtain at once finite unions. Note moreover that our definitions immediately give that λ is *finitely additive* on Σ, since given $A, B \in \Sigma$ disjoint,

$$\lambda(A \cup B) = \lambda((A \cup B) \cap A) + \lambda((A \cup B) \cap A^c),$$

which by disjointness unravels as

$$\lambda(A) + \lambda(B).$$

It remains to show closure under countable unions and countable intersections. However, given the previous work, this now reduces to showing closure under countable unions of *disjoint* sets in Σ.

Let $(A_n)_{n \in \mathbb{N}}$ be a sequence of disjoint sets in Σ. Let $A = \bigcup_{n \in \mathbb{N}} A_n$. Fix $B \subset S$. Note that since λ is monotone (in the sense, $C \subset C' \Rightarrow \lambda(C) \leq \lambda(C')$)) we have for any $N \in \mathbb{N}$

$$\lambda(B \cap A) \geq \lambda(B \cap \bigcup_{n \leq N} A_n).$$

Then any easy induction on N using the disjointness of the sets gives $\lambda(B \cap \bigcup_{n \leq N} A_n) = \sum_{n \leq N} \lambda(B \cap A_n)$. (For the inductive step: Use that since $A_N \in \Sigma$ we have $\lambda(B \cap \bigcup_{n \leq N} A_n) = \lambda((B \cap \bigcup_{n \leq N} A_n) \cap A_N) + \lambda((B \cap \bigcup_{n \leq N} A_n) \cap A_N^c) = \lambda(B \cap A_N) + \lambda(B \cap \bigcup_{n \leq N-1} A_n)$.) On the other hand, the assumption that λ is an outer measure give the inequality in the other way, and hence

$$\lambda(B \cap A) = \sum_{n \in \mathbb{N}} \lambda(B \cap A_n).$$

Finally, putting this altogether with the task at hand we have at every N

$$\lambda(B) = \lambda(B \cap \bigcup_{n \leq N} A_n) + \lambda(B \cap (\bigcup_{n \leq N} A_n)^c) \geq \lambda(B \cap \bigcup_{n \leq N} A_n) + \lambda(B \cap (\bigcup_{n \in \mathbb{N}} A_n)^c) = (\sum_{n \leq N} \lambda(B \cap A_n)) + \lambda(B \cap (\bigcup_{n \in \mathbb{N}} A_n)^c),$$

and thus taking the limit

$$\lambda(B) \geq (\sum_{n \in \mathbb{N}} \lambda(B \cap A_n)) + \lambda(B \cap (\bigcup_{n \in \mathbb{N}} A_n)^c) = \lambda(B \cap A) + \lambda(B \cap A^c);$$

by λ being an outer measure we get the inequality in the other direction and are done.

The argument from the last paragraph also shows that λ is countably additive on Σ, since in the equation

$$\lambda(B \cap A) = \sum_{n \in \mathbb{N}} \lambda(B \cap A_n)$$

we could as well have taken $B = S$. $\square$

This is good news, but doesn't yet solve the riddle of the existence of non-trivial measures: For all we know at this stage, Σ might typically end up as the two element σ-algebra $\{\emptyset, S\}$.

Theorem 3.3 *(Carathéodory extension theorem, part II) Let S be a set and let $\Sigma_0 \subset \mathcal{P}(S)$ be an algebra and let*

$$\mu_0 : \Sigma_0 \to \mathbb{R}^{\geq 0} \cup \{\infty\}$$

have $\mu_0(\emptyset) = 0$ and be σ-additive on its domain. (That is to say, if $(A_n)_{n \in \mathbb{N}}$ is a sequence of disjoint sets in Σ_0 and and if $\bigcup_{n \in \mathbb{N}} A_n \in \Sigma_0$, then $\mu_0(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu_0(A_n)$.)

Then μ_0 extends to a measure on the σ-algebra generated by Σ_0.

Proof Here the *σ-algebra generated by* Σ_0 means the smallest σ-algebra containing Σ_0 – it can be formally defined as the intersection of all σ-algebras containing Σ_0.

First of all we can apply the last theorem to obtain an outer measure λ, with $\lambda(A)$ being set equal to the infinum of all $\sum_{n \in \mathbb{N}} \mu_0(A_n)$ with each $A_n \in \Sigma_0$ and $A \subset \bigcup_{n \in \mathbb{N}} A_n$. The task which confronts us is to show that the σ-algebra indicated in 3.2 extends Σ_0 and that the measure λ extends the function μ_0. These are consequences of the next two claims.

Claim: For $A \in \Sigma_0$ and $B \subset S$

$$\lambda(B) = \lambda(B \cap A) + \lambda(B \cap A^c).$$

Proof of Claim: Clearly $\lambda(B) \leq \lambda(B \cap A) + \lambda(B \cap A^c)$. For the converse direction, suppose $(A_n)_{n \in \mathbb{N}}$ is a sequence of sets in Σ_0 with $B \subset \bigcup A_{n \in \mathbb{N}}$. Then at each n

$$\mu_0(A_n) = \mu_0(A_n \cap A) + \mu_0(A_n \cap A^c)$$

by the additivity properties of μ_0. Thus

$$\lambda(B \cap A) + \lambda(B \cap A^c) \leq \sum_{n \in \mathbb{N}} \mu_0(A \cap A_n) + \sum_{n \in \mathbb{N}} \mu_0(A^c \cap A_n) = \sum_{n \in \mathbb{N}} \mu_0(A_n).$$

(Claim$\square$)

Claim: $\mu_0(A) = \lambda(A)$ for any $A \in \Sigma_0$.

Proof of Claim: Suppose $(A_n)_{n \in \mathbb{N}}$ is a sequence of sets in Σ_0 with $A \subset \bigcup_{n \in \mathbb{N}} A_n$. We need to show that

$$\mu_0(A) \leq \sum_{n \in \mathbb{N}} \mu_0(A_n).$$

After replacing each A_n by $A_n \setminus \bigcup_{i<n} A_i$ we may assume the sets are disjoint. But consider $B_n = A_n \cap A$. $\mu_0(A) = \sum_{n \in \mathbb{N}} \mu_0(B_n)$ by the σ-additivity assumption on μ_0. Since σ-additivity implies finite additivity and hence that μ_0 is monotone, at each $n, \mu_0(B_n) \leq \mu_0(A_n)$. Thus

$$\mu_0(A) = \sum_{n \in \mathbb{N}} \mu_0(B_n) \leq \sum_{n \in \mathbb{N}} \mu_0(A_n).$$

(Claim$\square$)

$\square$

Now we are in a position to define Lebesgue measure rigorously.

Definition $B \subset \mathbb{R}^N$ is *Borel* if it appears in the smallest σ-algebra containing the open sets.

Of course one issue is why this is even well defined. Why should there be a *unique* smallest such algebra? The answer is that we can take the intersection of all σ-algebras containing the open sets and this is easily seen to itself be a σ-algebra.

Theorem 3.4 *There is a σ-additive measure m on the Borel subsets of $\mathbb{R}$ with*

$$m(\{x \in \mathbb{R} : a < x \leq b\}) = b - a$$

for all $a < b$ in $\mathbb{R}$.

Proof Let us call a $A \subset \mathbb{R}$ a *fingernail set* if it has the form

$$A = (a, b] =_{\mathrm{df}} \{x \in \mathbb{R} : a < x \leq b\}$$

for some $a < b$ in the extended real number line, $\{-\infty\} \cup \mathbb{R} \cup \{+\infty\}$. Let us take Σ_0 to be the collection of all finite unions of finger nail sets. It can be routinely checked that this is an algebra and every non-empty element of Σ_0 can be uniquely written in the form

$$(a_1, b_1] \cup (a_2, b_2] \cup ... \cup (a_n, b_n],$$

for some $n \in \mathbb{N}$ and $a_1 < b_1 < a_2 < ... < b_n$ in the extended real number line. For $A = (a_1, b_1] \cup (a_2, b_2] \cup ... \cup (a_n, b_n]$ we define

$$m_0(A) = (b_1 - a_1) + (b_2 - a_2) + ...(b_n - a_n).$$

Claim: If

$$(a, b] \subset \bigcup_{n \in \mathbb{N}} (a_n, b_n]$$

then $m_0((a, b]) \leq \sum_{n \in \mathbb{N}} m_0((a_n, b_n])$.

Proof of Claim: This amounts to show that if $(a, b] \subset \bigcup_{n \in \mathbb{N}} (a_n, b_n]$ then

$$\sum_{n \in \mathbb{N}} b_n - a_n \geq b - a.$$

Suppose instead $\sum_{n \in \mathbb{N}} b_n - a_n < b - a$. Choose $\epsilon > 0$ such that

$$\epsilon + \sum_{n \in \mathbb{N}} b_n - a_n < b - a.$$

Let $c_n = b_n + 2^{-n-1}\epsilon$. Then

$$\bigcup_{n \in \mathbb{N}} (a_n, c_n) \supset [a + \epsilon/2, b].$$

Applying Heine-Borel, as found at 2.11 , we can find some N such that

$$\bigcup_{n \leq N} (a_n, c_n) \supset [a + \epsilon/2, b].$$

The next subclaim states that after possibly changing the enumeration of the sequence $(a_n, c_n)_{n \leq N}$ we may assume that the intervals are consecutively arranged with overlapping end points.

Subclaim: we may assume without loss of generality that at each $i < N$ we have $a_{i+1} \leq c_i$.

Proof of Subclaim: First of all, since the index set is now finite, we may assume that no proper subset $Z \subsetneqq \{1, 2, ...N\}$ has $[a + \frac{\epsilon}{2}, b] \subset \bigcup_{n \in Z}(a_n, c_n)$. After possibly reordering the sequence we can assume $a_1 \leq a_j$ all $j \leq N$. Then we have $a_1 < a + \frac{\epsilon}{2}$ and since $\bigcup_{1<n\leq N}(a_n, c_n)$ does *not* include $[a + \frac{\epsilon}{2}, b]$ we have $c_1 > a + \frac{\epsilon}{2}$. Since $c_1 \in \bigcup_{n\leq N}(a_n, c_n)$ (there is another possibility, which is $c_1 > b$, but then $N = 1$ and the claim is trivialized) we have some j with $a_j < c_1 < c_j$. Without loss of generality $j = 2$, and then we continue so on. (□Subclaim)

Then

$$\sum_{i\in\mathbb{N}} b_i - a_i \geq \sum_{i\geq N} c_i - a_i \geq c_N - a_N + \sum_{i<N} a_{i+1} - a_i.$$

This is one of those telescoping sums where the middle terms all cancels out and we are left with

$$\sum_{i\geq N} c_i - a_i \geq c_N - a_1,$$

which turn must equal at least $b - a - \epsilon/2$, which contradicts our initial assumption of

$$\sum_{i\in\mathbb{N}} b_i - a_i < b - a - \frac{\epsilon}{2}.$$

(Claim□)

Claim: If $\{(a_n, b_n] : n \in \mathbb{N}\}$ is a *disjoint* sequence of fingernail sets with

$$\bigcup_{n\in\mathbb{N}} (a_n, b_n] \subset (a, b],$$

then

$$\sum_{n\in\mathbb{N}} m_0((a_n, b_n]) = \sum_{n\in\mathbb{N}} b_n - a_n \leq b - a.$$

Proof of Claim: It suffices to show that at each $N \in \mathbb{N}$ we have $\sum_{n\leq N} b_n - a_n \leq b - a$. After reordering we can assume that at any $i \leq j \leq N$ we have $a_i \leq a_j$. Then disjointness of the sequence gives $a_{j+1} \geq b_j$ at each $j \leq N$. Note also that our assumptions imply that each $a_i \geq a$ and $b_i \leq b$. Then it all unravels with

$$\sum_{n\leq N} b_n - a_n \leq b_N - a_N + \sum_{n<N} a_{n+1} - a_n$$

$$= b_N - a_1 \leq b - a.$$

(Claim□)

Claim: m_0 is σ-additive on Σ_0.

Proof of Claim: Let

$$A = (a_1, b_1] \cup (a_2, b_2] \cup ...(a_N, b_N]$$

be in Σ_0. Suppose $\{(c_n, d_n] : n \in \mathbb{N}\}$ are disjoint fingernail sets with

$$A = \bigcup_{n\in\mathbb{N}} (c_n, d_n].$$

At each $i \leq N$ and $n \in \mathbb{N}$ let $B_{n,i} = (c_n, d_n] \cap (a_i, b_i]$. The intersection of two fingernail sets is again a fingernail set, and thus each $B_{n,i}$ can be written in the form

$$B_{n,i} = (c_{n,i}, d_{n,i}].$$

The last two claims give that

$$\sum_{i \leq N} d_{n,i} - c_{n,i} = d_n - c_n$$

and that

$$b_i - a_i = \sum_{n \in \mathbb{N}} d_{n,i} - c_{n,i}.$$

Putting this altogether we have

$$\sum_{n \in \mathbb{N}} m_0((c_n, d_n]) = \sum_{n \in \mathbb{N}} d_n - c_n$$

$$= \sum_{n \in \mathbb{N}, i \leq N} d_{n,i} - c_{n,i} = \sum_{i \leq N} b_i - a_i = m_0(A).$$

(Claim$\square$)

Thus by 3.3 m_0 extends to a measure m which will be defined on the σ-algebra generated by Σ_0, which is the collection of Borel subsets of $\mathbb{R}$. $\square$

We used the fingernail sets $(a, b]$ because they neatly generate an algebra Σ_0. There would be no difference using different kinds of intervals.

Exercise Let m be the measure from 3.4.
(i) Show that for any $x \in \mathbb{R}$, $m(\{x\}) = 0$.
(ii) Conclude that $m([a, b]) = m((a, b)) = m([a, b)) = b - a$.
(iii) Show that if A is a countable subset of $\mathbb{R}$, then $m(A) = 0$.

A couple of remarks about the proof of 3.4. First of all, we have been rather stingy in our statement. The m from theorem is only defined on the Borel sets, but the proof of 3.2 and 3.3 gives that it is defined on a σ-algebra at least equal to the Borel sets; in fact it is a lot more, though for certain historical and conceptual reasons I am stating 3.4 simply for the Borel sets. Another remark about the proof is that we have only shown the theorem for one dimension, but it certainly makes sense to consider the case for higher dimensional euclidean space. Indeed one can prove that at every N there is a measure m_N on the Borel subsets of $\mathbb{R}^N$ such that for any rectangle of the form

$$A = (a_1, b_1] \times (a_2, b_2] \times ...(a_N, b_N]$$

we have

$$m_N(A) = (b_1 - a_1) \cdot (b_2 - a_2)...(b_N - a_N).$$

Theorem 3.5 *Let $N \in \mathbb{N}$ and $\Sigma \subset \mathcal{P}(\mathbb{R}^N)$ the σ-algebra of Borel subsets of N-dimensional Euclidean space. Then there is a measure*

$$m_N : \Sigma \to \mathbb{R}$$

such that whenever $A = I_1 \times I_2 \times ... \times I_N$ is a rectangle, each I_n an interval of the form (a_n, b_n), $[a_n, b_n)$, $(a_n, b_n]$, or $[a_n, b_n]$ we have

$$m_N(A) = (b_1 - a_1) \times (b_2 - a_2) \times ...(b_N - a_N).$$

A proof of this is given in [7]. In any case, the existence of such measures in higher dimension will follow from the one dimensional case and the section on product measures and Fubini's theorem later in the notes.

Finally, nothing has been said about uniqueness. One might in principle be concerned whether the measure m on the Borel sets in 3.4 has been properly defined – or whether there might be many such measures with the indicated properties. Although I did not pause to explicitly draw out this point, the measure indicated is unique for the reason that the measure of 3.3 is unique under modest assumptions. It is straightforward and I will leave it as an exercise.

Definition A measure μ on a measure space (X, Σ) is *σ-finite* if X can be written as a countable union of sets in Σ on which μ is finite.

Exercise Show that any measure m on $\mathbb{R}$ satisfying the conclusion of 3.4 is σ-finite.

Exercise (i) Let Σ_0 be an algebra, $\mu_0 : \Sigma_0 \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ σ-additive on its domain. Suppose S can be written as a countable union of sets in Σ_0 each of which has finite value under μ_0. Let $\Sigma \supset \Sigma_0$ be the σ-algebra generated by Σ_0 and let $\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ be a measure.

Then for every $A \in \Sigma$ we have

$$\mu(A) = \inf\{\sum_{n \in \mathbb{N}} \mu_0(A_n) : A \subset \bigcup_{n \in \mathbb{N}} A_n; \text{ each } A_n \in \Sigma_0\}.$$

(Hint: Write $S = \bigcup_{n \in \mathbb{N}} S_n$, each $S_n \in \Sigma_0$, each $\mu_0(S_n) < \infty$. It suffices to consider the case that $A \subset S_n$ for some n. By comparing A with its complement A^c and applying additivity, it suffices to show $\mu(A) \leq \inf\{\sum_{n \in \mathbb{N}} \mu_0(A_n) : A \subset \bigcup_{n \in \mathbb{N}} A_n; \text{ each } A_n \in \Sigma_0\}$ and $\mu(A^c) \leq \inf\{\sum_{n \in \mathbb{N}} \mu_0(A_n) : A^c \subset \bigcup_{n \in \mathbb{N}} A_n; \text{ each } A_n \in \Sigma_0\}$.)

(ii) Conclude that there is a *unique* measure satisfying 3.5.

Definition The unique measure described by the above theorem 3.5 is called *the Lebesgue measure* on $\mathbb{R}^N$.

Lemma 3.6 *Let $A \subset \mathbb{R}$ be Borel. Let m be Lebesgue measure on $\mathbb{R}$. Suppose $m(A) < \infty$. Let $\epsilon > 0$.*

(i) Then there exists an open set $O \subset \mathbb{R}$ such that

$$A \subset O$$

and

$$m(O) < m(A) + \epsilon.$$

(ii) And there exists a closed set $C \subset \mathbb{R}$ such that

$$C \subset A$$

and

$$m(A) < m(C) + \epsilon.$$

Proof (i) This is a consequence of our proof of 3.4. Implicitly we appealed to the existence of an outer measure via 3.3. This means here that for any A in the σ-algebra on which m is defined and for any $\epsilon > 0$ we have some sequence of fingernail sets, $(a_1, b_1], (a_2, b_2], ..$ with

$$A \subset \bigcup_{n \in \mathbb{N}} (a_n, b_n]$$

and

$$\sum_{n \in \mathbb{N}} b_n - a_n < m(A) + \epsilon.$$

Choose $\delta > 0$ with $\sum_{n\in\mathbb{N}} b_n - a_n < m(A) + \epsilon - \delta$. Let $c_n = b_n + \delta 2^{-n}$. We finish with

$$O = \bigcup_{n\in\mathbb{N}} (a_n, c_n).$$

(ii) It suffices to consider the case that $A \subset [a, b]$ for some $a < b$. Appealing to part (i), let $O \supset [a, b] \cap A^c$ be open with $m(O) < m([a, b] \cap A^c) + \epsilon$. Let $C = [a, b] \setminus O$. □

Formally I have just set up the Lebesgue measure on the Borel subsets of $\mathbb{R}$, but the proofs of 3.2 and 3.3 suggest that perhaps it can be sensibly defined on a rather larger σ-algebra.

Definition Let m^* be the outer measure used to define Lebesgue measure – in that

$$m^*(A)$$

equals the infinum of

$$\{\sum_{i\in\mathbb{N}}(b_i - a_i) : ((a_i, b_i])_{i\in\mathbb{N}} \text{ is a sequence of fingernail sets with } A \subset \bigcup_{i\in\mathbb{N}}(a_i, b_i]\}.$$

A subset $B \subset \mathbb{R}$ is *Lebesgue measurable* if for every $A \subset \mathbb{R}$ we have $m^*(A) = m^*(A \cap B) + m^*(A \cap B^c)$.

From theorems 3.2 and 3.3 we obtain that the Lebesgue measurable sets in $\mathbb{R}$ form a σ-algebra and m extends to a measure on that σ-algebra. In a similar vein:

Exercise Show that if $B \subset \mathbb{R}$ is Lebesgue measurable then there are Borel $A_1, A_2 \subset \mathbb{R}$ with

$$A_1 \subset B \subset A_2$$

and $m(A_2 \setminus A_1) = 0$. (Hint: This follows from the proof of 3.6.)

One may initially wonder whether the Lebesgue measurable sets are larger than the Borel.

The short answer is that not only are there more, there are vastly more. Take a version of the Cantor set with measure zero. For instance, the standard construction where we remove the interval $(1/4, 3/4)$ from $[0, 1]$, then $(1/16, 3/16)$ from $[0, 1/4]$ and $(13/16, 15/16)$ from $[3/4, 1]$, and continute, iteratively removing the middle halves. The final result will be a closed, nowhere dense set. A routine compactness argument shows that it is non-empty and has no isolated points. With a little bit more work we can show it is actually homeomorphic to $\prod_{\mathbb{N}}\{0, 1\}$, and hence has size $2^{\aleph_0}$.

Its Lebesgue measure is zero, since the set remaining after n many steps has measure 2^{-n}. Any subset of a Lebesgue measurable set *of measure zero* is again Lebesgue measurable, thus all its subsets are Lebesgue measurable. Since it has $2^{(2^{\aleph_0})}$ many subsets, we obtain $2^{(2^{\aleph_0})}$ Lebesgue measurable sets. On the other hand it can be shown (see for instance [6]) that there are only $2^{\aleph_0}$ many Borel sets – and thus not every Lebesgue measurable set is Borel.

Fine. But then of course it is natural to be curious about the other extreme. In fact, there exist subsets of $\mathbb{R}$ which are not Lebesgue measurable.

Lemma 3.7 *There exists a subset $V \subset [0, 1]$ which is not Lebesgue measurable.*

Proof For each $x \in [0, 1]$ let Q_x be $\mathbb{Q} + x \cap [0, 1]$ – that is to say, the set of $y \in [0, 1]$ such that $x - y \in \mathbb{Q}$. Note that $Q_x = Q_z$ if and only if $x - z \in \mathbb{Q}$.

Now let $V \subset [0, 1]$ be a set which intersects each Q_x exactly once. Thus for each $x \in [0, 1]$ there will be exactly one $z \in V$ with $x - z \in \mathbb{Q}$.

I claim V is not Lebesgue measurable.

For a contradiction assume it is. Note then that for any $q \in \mathbb{Q}$ the translation $V + q = \{z + q : z \in V\}$ has the same Lebesgue as V. This is simply because the entire definition of Lebesgue measure was translation invariant.

The first case is that V is null. But then $\bigcup_{q \in \mathbb{Q}} V + q$ is a countable union of null sets covering $[0, 1]$, with a contradiction to $m([0, 1]) = 1$.

Alternatively, if $m(V) = \epsilon > 0$, then $\bigcup_{q \in \mathbb{Q}, 0 \leq q \leq 1} V + q$ is an infinite union of disjoint sets all of measure ϵ, all included in $[-1, 2]$, with a contradiction to $m([-1, 2]) = 3 < \infty$. $\square$

There is something curious about the construction of the set V above. It is created by an appeal to the *axiom of choice* – we choose exactly one point from each set of the form Q_x, but the final set V is not itself presented with any easy description.

One might then ask if there is an *explicit* or *concrete* example of a set which is not Lebesgue measurable, or alternatively whether one can prove the existence of sets which are not Lebesgue measurable without appealing to the axiom of choice. It takes some care to make these questions mathematically precise, but the answer to both ultimately is in the negative – see [11].

Bear in mind that there are other kinds of spaces and objects to which we might wish to assign something like a measure.

Examples (i) For a finite set X and $\mathcal{P}(X)$ the set of all subsets of X, we could take the counting measure: $\mu(A) = |A|$, the size of A.

(ii) For $\{H, T\}^N$, which could be thought of as tossing a coin with outcome either "H" (heads) or "T" (tails), we could take the normalized counting measure:

$$\mu(A) = \frac{|A|}{2^N}.$$

(iii) A natural variation on (ii) is to take the *infinite* product of the coin tossing measure. Let

$$\prod_{\mathbb{N}} \{H, T\}$$

be the collection of all functions $f : \mathbb{N} \to \{H, T\}$. For $i_1, i_2, ..., i_N$ distinct elements of $\mathbb{N}$ and $S_1, S_2, ..., S_N \in \{H, T\}$ let

$$\mu(\{f \in \prod_N \{H, T\} : f(i_1) = S_1, f(i_2) = S_2, ..., f(i_N) = S_N\} = 2^{-N}.$$

3.3 shows that μ extends to a measure on the subsets of $\prod_{\mathbb{N}} \{H, T\}$ which are Borel with respect to the product topology.

(iv) Another way in which (ii) can be altered is to adjust the measure on $\{H, T\}$. We might instead be working with a slightly biased coin, that comes down heads 7 times out of ten. Then for each $f : \{1, 2, ..., N\} \to \{H, T\}$ we set

$$\mu(\{f\}) = (\frac{7}{10})^{|\{i : f(i) = H\}|} \cdot (\frac{3}{10})^{|\{i : f(i) = T\}|}.$$

Similarly we could define this lopsided measure on the space of infinite runs.

(v) Somewhat more loosely, imagine we are dealing with some experiment, such as shooting gamma rays into a metal alloy, and X is the space of all possible outcomes to the experiment. Let $f : X \to \mathbb{R}$ be some function which arises from measuring some property of the outcome (in this case the heat of the metal alloy). We could then define a measure on $\mathbb{R}$ by letting $\mu(A)$ be the *probability* that f assumes its value in A.

We have already fought a considerable battle simply to show that interesting measures, such as Lebesgue measure, indeed exist. From now on we will take this as a given, and consider more the abstract properties of measures. Here there is the key notion of *completion* of a measure.

Definition Let (X, Σ) be a measure space and $\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ a measure. $M \subset X$ is said to be *measurable with respect to* μ if there are $A, B \in \Sigma$ with $A \subset M \subset B$ and

$$\mu(B \setminus A) = 0.$$

Beware: Often authors simply write "measurable" instead of "measurable with respect to μ" when context makes the intended measure clear.

Lemma 3.8 *Let μ be a measure on the measure space (X, Σ). Then the measurable sets form a σ-algebra*

Proof First suppose M is measurable as witnessed by $A, B \in \Sigma$, $A \subset M \subset B$. Then $A^c = X \setminus A, B^c = X \setminus B$ are in Σ and have $A^c \supset M^c \supset B^c$. Since $A^c \setminus B^c = B \setminus A$, this witnesses M^c measurable.

If $(M_n)_{n\in\mathbb{N}}$ is a sequence of measurable sets, as witnessed by $(A_n)_{n\in\mathbb{N}}, (B_n)_{n\in\mathbb{N}}$, with $A_n \subset M_n \subset B_n$, then

$$\bigcup_{n\in\mathbb{N}} A_n \subset \bigcup_{n\in\mathbb{N}} M_n \subset \bigcup_{n\in\mathbb{N}} B_n.$$

On the other hand

$$\bigcup_{n\in\mathbb{N}} B_n \setminus \bigcup_{n\in\mathbb{N}} A_n \subset \bigcup_{n\in\mathbb{N}} (B_n \setminus A_n),$$

and so $\bigcup_{n\in\mathbb{N}} B_n \setminus \bigcup_{n\in\mathbb{N}} A_n$ is null, as required to witness $\bigcup_{n\in\mathbb{N}} M_n$ null. □

Definition Let μ be a measure on the measure space (X, Σ). For M measurable, as witnessed by $A \subset M \subset B$ with $A, B \in \Sigma$, $\mu(B \setminus A) = 0$, we let $\mu^*(M) = \mu(A)(= \mu(B))$. We call μ^* the *completion of* μ.

As with Lebesgue measure, we will frequently slip in to the minor logical sin of using the same symbol for μ as its extension μ^* to the measurable sets. A more serious issue is to check the measure is well defined.

Lemma 3.9 *The completion of μ to the measurable sets is well defined.*

Proof If $A_1 \subset M \subset B_1$ and $A_2 \subset M \subset B_2$ both witness M measurable, then

$$A_1 \Delta A_2 \subset M \setminus A_1 \cup M \setminus A_2 \subset B_1 \setminus A_1 \cup B_2 \setminus A_2.$$

and hence is null. □

Lemma 3.10 *Let μ be a measure on a measure space (X, Σ). Let Σ^* be the σ-algebra of measurable sets. Then the completion defined above,*

$$\mu^* : \Sigma^* \to \mathbb{R}^{\geq 0} \cup \{\infty\},$$

is a measure.

Proof Exercise. □

Exercise For $A \subset \mathbb{N}$ let

$$\mu(A) = |A|$$

in the event A is finite, and equal ∞ otherwise. Show that μ is a measure on the σ-algebra $\mathcal{P}(\mathbb{N})$.

Exercise Let $f : \mathbb{R} \to \mathbb{R}$ be non-decreasing, in the sense that

$$a \leq b \Rightarrow f(a) \leq f(b).$$

Show that the image of f, $f[\mathbb{R}]$, is Borel.

Exercise Let

$$\prod_{\mathbb{N}} \{H, T\}$$

be the collection of all functions $f : \mathbb{N} \to \{H, T\}$. equipped with the product topology. For $\vec{i} = i_1, i_2, ..., i_N$ distinct elements of $\mathbb{N}$ and $\vec{S} = S_1, S_2, ..., S_N \in \{H, T\}$ let

$$A_{\vec{i},\vec{S}} = \{f \in \prod_{\mathbb{N}} \{H, T\} : f(i_1) = S_1, f(i_2) = S_2, ..., f(i_N) = S_N\}.$$

(i) Show that the sets of the form $A_{\vec{i},\vec{S}}$ form an algebra (i.e. closed under *finite* unions, intersections, and complements).

(ii) Show that

$$\mu_0(\{f \in \prod_{N} \{H, T\} : f(i_1) = S_1, f(i_2) = S_2, ..., f(i_N) = S_N\}) = 2^{-N}$$

defines a function which is σ-additive on its domain.

(iii) Show that μ_0 extends to a measure μ on the Borel subsets of $\prod_{\mathbb{N}} \{H, T\}$.

(iv) At each N let A_N be the set of $f \in \prod_{\mathbb{N}} \{H, T\}$

$$|\{n < N : f(n) = H\}| < \frac{N}{3}.$$

Show that $\mu(A_N) \to 0$ as $N \to \infty$.

(v) Let A be the set of f such that there exist infinitely many N with $f \in A_N$. Show that A is Borel.

Exercise Let $A \subset \mathbb{R}$ be Lebesgue measurable. Show that $m(A)$ is the supremum of

$$\{m(K) : K \subset A, K \text{ compact}\}.$$

Exercise Show that if we successively remove middle thirds from $[0, 1]$, and then from $[0, 1/3]$ and $[2/3, 1]$, and then from $[1.1/9]$, $[2/9, 1/3]$, $[2/3, 7/9]$, and $[8/9, 1]$, and so on, then the set at the end stage of this construction has measure zero.

The resulting set is called *the Cantor set*, and is a source of examples and counterexamples in real analysis – given that it is closed, nowhere dense, and without isolated points. More generally a set formed in a similar way is often called *a Cantor set*.

Exercise Show that if we adjust the process by removing a middle tenths, then we again end up with a Cantor set having measure zero.

Exercise Show that if we instead remove the middle tenth, and at the next step the middle one hundredths, and then at the next step the middle one thousands, and so on, then we end up with a Cantor set which has positive measure.

4 The general notion of integration and measurable function

We will give a rigorous foundation to Lebesgue integration as well as integration on general measure spaces. Since the notion of integration is so closely intertwined with linear notions of adding or subtracting, I will first give the definitions of the linear operations for functions.

Definition Let X be a set and $f, g : X \to \mathbb{R}$. Then we define the functions $f + g$ and $f - g$ by

$$f + g : X \to \mathbb{R},$$
$$x \mapsto f(x) + g(x),$$

and

$$f - g : X \to \mathbb{R},$$
$$x \mapsto f(x) - g(x),$$

and

$$f \cdot g : X \to \mathbb{R},$$
$$x \mapsto f(x)g(x).$$

Similarly, for $c \in \mathbb{R}$ we define

$$cf : X \to \mathbb{R},$$
$$x \mapsto cf(x)$$

and

$$c + f : X \to \mathbb{R},$$
$$x \mapsto c + f(x).$$

Definition Let (X, Σ) be a measure space equipped with a measure

$$\mu : \Sigma \to \mathbb{R}.$$

A function $f : X \to \mathbb{R}$ is *measurable* if for any open set $U \subset \mathbb{R}$

$$f^{-1}[U] \in \Sigma.$$

A function

$$h : X \to \mathbb{R}$$

is *simple* if we can partition X into finitely many measurable sets $A_1, A_2, ..., A_n$ with h assuming a constant value a_i on each A_i. If $\mu(A_i) < \infty$ whenever $a_i \neq 0$ we say that h is integrable and let

$$\int_X h d\mu$$

be the sum of all $\mu(A_i) \cdot a_i$ for $a_i \neq 0$. For $B \subset X$ a measurable set, we define

$$\int_B h d\mu$$

to be the sum of $\mu(A_i \cap B) \cdot a_i$ for $a_i \neq 0$.

Exercise Show that every simple function is measurable.

Definition For (X, Σ, μ) as above, $f : X \to \mathbb{R}$ measurable and non-negative ($f(x) \geq 0$ all $x \in X$), we let

$$\int_X f \, d\mu = \sup\{\int_X h \, d\mu : \ h \text{ is simple and } 0 \leq h \leq f\}.$$

In general given $f : X \to \mathbb{R}$ measurable we can uniquely write $f = f^+ - f^-$ where f^+, f^- are both non-negative and have disjoint supports. Assuming

$$\int_X f^+ d\mu, \int_X f^- d\mu$$

are both finite we say that f is *integrable* and let

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu.$$

We have implicitly used above that f^+, f^- will be measurable. This is easy to check. You might also want to check that we could have instead used the definition $f^+ = \frac{1}{2}(|f| + f)$, and $f^- = \frac{1}{2}(|f| - f)$.

Definition For (X, Σ, μ) as above, $f : X \to \mathbb{R}$ measurable and $B \in \Sigma$,

$$\int_B f d\mu = \int_X \chi_B \cdot f d\mu.$$

Exercise We could alternatively have defined $\int_B f d\mu$ to be the supremum of

$$\int_B h d\mu$$

for h ranging over simple functions with $h \leq f$. Show this definition is equivalent to the one above.

Lemma 4.1 *Let (X, Σ) be a measure space, μ a measure defined on X. Let $f : X \to \mathbb{R}$ be a simple integrable function. Let $c \in \mathbb{R}$. Then*

$$\int_X c f d\mu = c \int_X f d\mu.$$

Proof Exercise. $\square$

Lemma 4.2 *Let (X, Σ) be a measure space, μ a measure defined on X. Let $f, g : X \to \mathbb{R}$ be simple integrable functions with $f(x) \leq g(x)$ at all x. Then*

$$\int_X f d\mu \leq \int_X g d\mu.$$

Proof Exercise. $\square$

Lemma 4.3 *Let X, Σ, μ be as above. Let f_1, f_2 be simple integrable functions. Then*

$$\int_X (f_1 + f_2) d\mu = \int_X f_1 d\mu + \int_X f_2 d\mu.$$

Proof Suppose $f_1 = \sum_{i \le N} a_i \chi_{A_i}$, $f_2 = \sum_{i \le M} b_i \chi_{B_i}$. Then at each $i \le N, j \le M$ let $C_{i,j} = A_i \cap B_j$.

$$f_1 + f_2 = \sum_{i \le N, j \le M} (a_i + b_j)\chi_{i,j}.$$

Since $\mu(A_i) = \sum_{j \le M} \mu(C_{i,j})$ and $\mu(B_j) = \sum_{i \le N} \mu(C_{i,j})$ we have

$$\int_X f_1 d\mu = \sum_{i \le N, j \le M} a_i \mu(C_{i,j})$$

and

$$\int_X f_1 d\mu = \sum_{i \le N, j \le M} b_j \mu(C_{i,j}).$$

Thus

$$\int_X f_1 d\mu + \int_X f_2 d\mu = \sum_{i \le N, j \le M} (a_i + b_j)\mu(C_{i,j}),$$

which in turn equals $\int_X (f_1 + f_2) d\mu$. □

Definition Let S be a set. A *partition* of S is a collection $\{S_i : i \in I\}$ such that:

1. each $S_i \subset S$;
2. $S = \bigcup_{i \in I} S_i$;
3. for $i \neq j$ we have $S_i \cap S_j = \emptyset$.

In other words, a partition is a division of the set into disjoint subsets.

Lemma 4.4 *Let X, Σ, μ be as above. Let $f : X \to \mathbb{R}$ be integrable. Let $(A_i)_{i \in \mathbb{N}}$ be a partition of X into countably many sets in Σ. Then*

$$\int_X f d\mu = \sum_{i \in \mathbb{N}} \int_{A_i} f d\mu$$

.

Proof Wlog $f \ge 0$.

First to see that $\int_X f d\mu \ge \sum_{i \in \mathbb{N}} \int_{A_i} f d\mu$, suppose $h_i \le f \cdot \chi_{A_i}$ at each $i \le N$. Then $\sum_{i \le N} h_i \le f$ and $\sum \int h_i d\mu = \int \sum h_i d\mu$.

Conversely, if $h \le f$ is simple, write it as

$$h = \sum_{j \le k} a_j \chi_{B_j}$$

with each $a_j > 0$, which implies each $\mu(B_j) < \infty$ by integrability of f. Consider some $\epsilon > 0$. Go to a large enough N with

$$\mu(\bigcup_{j \le k} B_j \setminus \bigcup_{i > N} A_i) < \frac{\epsilon}{\sum_j a_j}.$$

Then

$$\sum_{i \le N} \int_{A_i} f d\mu > \int_X h d\mu - \mu(\bigcup_{j \le k} B_j \setminus \bigcup_{i > N} A_i) \sum_j a_j > \int h d\mu - \epsilon.$$

Letting $\epsilon \to 0$ finishes the proof. □

Lemma 4.5 *Let X, Σ, μ be as above. Let $f : X \to \mathbb{R}$ be integrable. Then for each $\epsilon > 0$ we can find a set B with $\mu(B)$ finite and $\int_{X \setminus B} f d\mu < \epsilon$.*

Proof Wlog, $f \geq 0$. At each $N \in \mathbb{N}$ let $A_N = \{x : \frac{1}{N} \leq f(x) < \frac{1}{N-1}\}$. (Note then that $f^{-1}([1, \infty)) = A_1$). The measure of each A_N is finite since $\int_{A_N} f d\mu \geq \mu(A_N)\frac{1}{N}$. By 4.4, we have some M with $\int_{\bigcup_{N \geq M}(A_N)} f d\mu < \epsilon$. □

Lemma 4.6 *Let (X, Σ) be a measure space. Let $\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ be a measure. Let $f : X \to \mathbb{R}$ be integrable. Then there is a sequence of simple functions, $(f_n)_{n \in \mathbb{N}}$ such that:*

1. *for a.e. $x \in X$, $f_n(x) \to f(x)$;*
2. *$|f_n(x)| < |f(x)|$ all $x \in X$;*
3. *at each $n \in \mathbb{N}, x \in X$, $f_n(x) \geq 0$ if and only if $f(x) \geq 0$.*

As a word on notation, we say that something happens "a.e." or "μ-a.e" if it is true off of some null sets – where a null set is some set in Σ whose value under μ is zero. Thus, "for a.e. $x \in X$, $f_n(x) \to f(x)$" means that there is some $B \in \Sigma$ with $\mu(B) = 0$ and $f_n(x) \to f(x)$ all $x \in B^c$.

Proof It suffices to consider the case of $f(x) \geq 0$ all $x \in X$. At each x and n let $k_n(x)$ be the largest k such that

$$\frac{k}{n!} \leq f(x).$$

Then let

$$f_n(x) = \min\{\frac{k_n(x)}{n!}, n\}.$$

It follows routinely from f being measurable that each k_n is measurable, and then that each f_n is measurable. Since the f_n's are measurable and finite valued, they are simple. Unwinding the definitions, we have for all $n \geq f(x)$ that

$$f_n(x) \geq f(x) - \frac{1}{n!}.$$

□

Lemma 4.7 *Let (X, Σ, μ) and $f : X \to \mathbb{R}$ be as in the last lemma. Let $(f_n)_{n \in \mathbb{N}}$ be as in the conclusion – that is to say,*

1. *each f_n is simple;*
2. *for a.e. $x \in X$, $f_n(x) \to f(x)$;*
3. *$|f_n(x)| < |f(x)|$ all $x \in X$;*
4. *at each $n \in \mathbb{N}, x \in X$, $f_n(x) \geq 0$ if and only if $f(x) \geq 0$.*

Then

$$\int_X f_n d\mu \to \int f d\mu.$$

Remark: Please note, the point of this lemma is *not* to say that we can have all the indicated properties along with $\int f_n \to \int f$. We already showed that in the proof above. The point is rather that once we have 1-4, *then* $\int f_n \to \int f$ follows automatically.

Proof Let the f_n's be as indicated. Again we can assume $f(x) \geq 0$ at all $x \in X$. Suppose we instead have some simple function

$$h = \sum_{i \leq N} a_i \cdot \chi_{B_i}$$

with

$$h(x) \leq f(x)$$

all x and at all n

$$\int h d\mu < \epsilon + \int f_n d\mu$$

for some fixed positive ϵ. We can assume that $a_i \neq 0$ at each $i \leq N$. Then it follows that $\bigcup_{i \leq N} B_i$ has finite measure. We may also assume each $b_i \geq 0$.

We want to show that $\int h d\mu \leq \lim \int f_n d\mu$ Let

$$\delta = \frac{\epsilon}{2(a_1 + a_2 + ... + a_N + \mu(B_1) + \mu(B_2) + ... + \mu(B_N))}.$$

For each n let D_n be the set of $x \in X$ such that

$$|f_n(x) - f(x)| < \delta.$$

Here $\bigcup_{n \in N} D_n$ is conull, $D_n \subset D_{n+1}$, and each D_n is in Σ. Thus we can find some D_k such that

$$\mu(\bigcup_{i \leq N} B_i - D_k) < \delta.$$

Then

$$h(x) \leq (f_k(x) + \delta)\chi_{D_k} + (a_1 + a_2 + ...a_N)\chi_{(B_1 \cup B_2 \cup ... B_N) \setminus D_k}.$$

Thus by 4.2 and 4.3 we have

$$\int_X h d\mu \leq \int_{D_k \cap (B_1 \cup B_2 \cup ... B_N)} (f_k + \delta) d\mu + \int_{(B_1 \cup B_2 \cup ... B_N) \setminus D_k} (a_1 + a_2 + ...a_N) d\mu$$

$$\leq \int_{D_k \cap (B_1 \cup B_2 \cup ... B_N)} f_k d\mu + \int_{D_k \cap (B_1 \cup B_2 \cup ... B_N)} \delta d\mu + \int_{(B_1 \cup B_2 \cup ... B_N) \setminus D_k} (a_1 + a_2 + ...a_N) d\mu$$

$$< \int_X f_k d\mu + \delta \cdot \mu(B_1 \cup B_2 \cup ...B_N) + (a_1 + a_2 + ...a_N)\mu((B_1 \cup B_2 \cup ...B_N) \setminus D_k) < \int f_k + \epsilon,$$

as required. □

Lemma 4.8 *Let (X, Σ) be a measure space, μ a measure defined on X. Let $f : X \to \mathbb{R}$ be an integrable function. Let $c \in \mathbb{R}$. Then*

$$\int_X cf d\mu = c \int_X f d\mu.$$

Proof Exercise. □

Lemma 4.9 *Let X, Σ, μ be as above. Let f_1, f_2 be integrable functions. Then*

$$\int_X (f_1 + f_2) d\mu = \int_X f_1 d\mu + \int_X f_2 d\mu.$$

Proof We want to put this into the framework of 4.7 – ideally choosing simple functions $g_1, g_2..., h_1, h_2, ...$ such that

1. at each x and n, $|g_n(x)| \leq |f_1(x)|$, and
2. at each x and n, $|h_n(x)| \leq |f_2(x)|$, and
3. $g_n(x) \to f_1(x)$ for a.e. x, and
4. $h_n(x) \to f_2(x)$ for a.e. x,

and then concluding that $g_n + h_n$ converge to $f_1 + f_2$ with $|g_n(x) + h_n(x)| \leq |f_1(x) + f_2(x)|$.

This will be fine if f_1 and f_2 are either both positive or both negative. The problem is that if they have different sign – for instance, if $f_1(x) = -6$, $f_2(x) = 5$, and we had unluckily chosen $g_n(x) = -4$, $h_n(x) = 2$.

Here is the solution to that minor technical issue. Let

$$B_1 = \{x : f_1(x), f_2(x) \geq 0\},$$
$$B_2 = \{x : f_1(x), f_2(x) < 0\},$$
$$B_3 = \{x : f_1(x) \geq 0, f_2(x) < 0, |f_1(x)| > |f_2(x)|\},$$
$$B_4 = \{x : f_1(x) \geq 0, f_2(x) < 0, |f_1(x)| \leq |f_2(x)|\},$$
$$B_5 = \{x : f_2(x) \geq 0, f_1(x) < 0, |f_1(x)| > |f_2(x)|\},$$
$$B_6 = \{x : f_2(x) \geq 0, f_1(x) < 0, |f_1(x)| \leq |f_2(x)|\}.$$

It suffices to show that on each B_i we have $\int_{B_i} f_1 d\mu + \int_{B_i} f_2 d\mu = \int_{B_i} (f_1 + f_2) d\mu$.

The sets B_1, B_2 are handled by the argument given above; I will skip the details. It is B_i for $i \geq 3$ which requires more work. All these cases are much the same, and so I will simply do B_3. First choose simple h_n's with $h_n(x) \to f_2(x)$ and $|h_n(x)| \leq |h_{n+1}(x)| \leq |f_2(x)|$ for a.e. $x \in B_3$. Now choose g_n on B_3 such that $g_n(x) \to f_1(x)$ and $|g_n(x)| \leq |g_{n+1}(x)| \leq |f_1(x)|$ *and* $|g_n(x)| \geq |h_n(x)|$; the last point is easily arranged since we can always replace g_n with $\max\{g_n(x), |h_n(x)|\}$. Then we have

1. at each x and n, $|g_n(x)| \leq |f_1(x)|$, and
2. at each x and n, $|h_n(x)| \leq |f_2(x)|$, and
3. $g_n(x) \to f_1(x)$ for a.e. x, and
4. $h_n(x) \to f_2(x)$ for a.e. x, and
5. $|g_n(x) + h_n(x)| \leq |f_1(x) + f_2(x)|$.

Apply 4.7 to f_1 and the g_n's we get $\int_{B_i} g_n d\mu \to \int_{B_i} f_1 d\mu$; to f_2 and the h_n's, $\int_{B_i} h_n d\mu \to \int_{B_i} f_2 d\mu$; finally, $\int_{B_i} (g_n + h_n) d\mu \to \int_{B_i} (f_1 + f_2) d\mu$ and for simple functions we already know that $\int_{B_i} (g_n + h_n) d\mu = \int_{B_i} g_n d\mu + \int_{B_i} h_n d\mu$. □

Lemma 4.10 *Let $C \subset O \subset \mathbb{R}$ with C closed and O open. Show that there is a continuous function $f : \mathbb{R} \to [0, 1]$ with $f(x) = 1$ at every point on C and $f(x) = 0$ at every point outside O.*

Proof Assume C is non empty and $O \neq \mathbb{R}$, or else the task is somewhat trivialized. For any set $A \subset \mathbb{R}$ let $d(x, A) = \inf\{|x - a| : a \in A\}$ – this is a continuous function in x.

Then let

$$f(x) = \min\{1, \frac{d(x, O^c)}{d(x, C)}\}.$$

□

Lemma 4.11 *Let h be a simple function on $\mathbb{R}$. Suppose h is integrable. Let $\epsilon > 0$. Then there is a continuous function*

$$f : \mathbb{R} \to \mathbb{R}$$

such that if we let $g(x) = |h(x) - f(x)|$ then

$$\int_{\mathbb{R}} gdm < \epsilon.$$

(Technically, when saying h is simple we should specify the corresponding σ-algebra on $\mathbb{R}$. The convention is to default to the Borel sets – thus, we intend that there be a partition on $\mathbb{R}$ into finitely many Borel sets and h is constant on each element of the partition.)

Proof It suffices to prove this in the case when h is the characteristic function of a single Borel set. But then this follows from 3.6 and 4.10. □

Corollary 4.12 *Let $h : \mathbb{R} \to \mathbb{R}$ be integrable. Let $\epsilon > 0$. Then there is a continuous function*

$$f : \mathbb{R} \to \mathbb{R}$$

such that if we let $g(x) = |h(x) - f(x)|$ then

$$\int_{\mathbb{R}} gdm < \epsilon.$$

Frequently we will want to integrate compositions of functions, just as in the last exercise. Here there is some very specific notation used in this context to guide us.

Notation Given some expression involving various variables, $x, y, z...$ and various functions, say $G(x, y, z, ...)$, the expression

$$\int_X G(x, y, ...)d\mu(x)$$

indicates that we are integrating the function

$$x \mapsto G(x, y, z, ...)$$

against the measure μ (and keeping $y, z, ...$ as fixed but possibly unknown quantities).

Thus in the exercise just above, for $g(x) = |f(x) - h(x)|$, instead of writing

$$\int_{\mathbb{R}} gdm < \epsilon$$

we could just as easily written

$$\int_R |f(x) - h(x)|d\mu(x).$$

The definition of integration can be extended to other settings.

Definition Let (X, Σ, μ) be a measure space equipped with a measure μ; we say that a function from X to $\mathbb{C}$ is *measurable* if the pullback of any open set in $\mathbb{C}$ is measurable.

For $f : X \to \mathbb{C}$ measurable, we write $f = \text{Re}f + i\text{Im}f$, where $\text{Re}f : X \to \mathbb{R}$ and $\text{Im}f : X \to \mathbb{R}$ are the real and imaginary parts. We say that f is *integrable* if both these functions are integrable in our earlier sense and let

$$\int_X fd\mu = \int_X \text{Re}fd\mu + i\int_X \text{Im}fd\mu.$$

In fact, it does not stop there. Given a suitable *linear* space $\mathbb{B}$ we can define integrals for suitably bounded functions $f : X \to \mathbb{B}$. In general terms, if the space $\mathbb{B}$ allows us to form finite sums and averages, then it makes sense to define integrals on $\mathbb{B}$-valued functions.

5 Convergence theorems

The order of presentation is following [7], but I am going to present the proofs without making any reference to $L^1(\mu)$ or the theory of Banach spaces. Eventually we will have to engage with these concepts, but not just yet.

Lemma 5.1 *Let (X, Σ) be a measure space. Let $\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ be a measure. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions, each $f_n : X \to \mathbb{R}$ integrable. Suppose each $f_n \geq 0$ and*

$$\sum_{n=1}^{n=\infty} \int f_n d\mu < \infty.$$

Then for μ-a.e. $x \in X$,

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

is well defined, and moreover

$$\int f d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

Proof Let B be the set of $x \in X$ such that the partial sums $\sum_{n=1}^{N} f_n(x)$ are unbounded. It is routine to show that this set is in Σ. I claim it is null.

Claim: $\mu(B) = 0$.

Proof of Claim: Suppose instead that $\mu(B) > 0$. Then at each $c > 0$ and $N \in \mathbb{N}$ let $B_{c,N} = \{x : \sum_{n=1}^{N} f_n(x) > c\}$. $B_{c,N} \subset B_{c,N+1}$ and

$$\bigcup_{N\in\mathbb{N}} B_{c,N} = B.$$

Thus there exists an N with $\mu(B_{c,N}) > \frac{1}{2}\mu(B)$. Then we obtain

$$\sum_{n=1}^{N} \int_{B_{c,N}} f_n d\mu = \int_{B_{c,N}} \sum_{n=1}^{N} f_n d\mu > c\frac{1}{2}\mu(B).$$

Since $c > 0$ was arbitrary, we have contradicted $\sum_{n=1}^{n=\infty} \int f_n d\mu < \infty$. (Claim$\square$)

Now define f on $X \setminus B$ as $f(x) = \sum f_n(x)$ (and set $f \equiv 0$ on B – though in terms of calculating the integral, the value of f on a null set is irrelevant).

Claim: $\int_X f d\mu \geq \sum_{n=1}^{\infty} \int_X f_n d\mu$.

Proof of Claim: At each $N \in \mathbb{N}$ we have

$$\int_X f d\mu \geq \int_X \sum_{n\leq N} f_n d\mu = \sum_{n\leq N} \int_X f_n d\mu.$$

(Claim$\square$)

Claim: $\int_X f d\mu \leq \sum_{n=1}^{\infty} \int_X f_n d\mu$.

Let $h \leq f$ be simple. Wlog $h \geq 0$. Write h as

$$\sum_{i \leq L} a_i \chi_{A_i},$$

where each $a_i > 0$. Let

$$C_N = \{x \in B : \forall M \geq N |\sum_{n=1}^{M} f_n(x) - f(x)| < \frac{\epsilon}{2\sum \mu(A_i)}\}.$$

Again the C_N's are increasing and their union is conull. Fix $\epsilon > 0$.

Subclaim: There is some N with

$$\int_{X \setminus C_N} h d\mu < \frac{\epsilon}{2}.$$

Proof of Subclaim: Let $D_n = C_n \setminus (\bigcup_{i<n} D_i)$. Then

$$\int_X h d\mu = \sum_{n \in \mathbb{N}} \int_{D_n} h d\mu$$

by 4.4. Since the integral is finite, we can find some N with

$$\sum_{n>N} \int_{D_n} h d\mu < \frac{\epsilon}{2}.$$

(Proof of Subclam□)

Then

$$\int_X h d\mu = \int_{\bigcup_{i \leq L} A_i} h d\mu < \frac{\epsilon}{2} + \int_{C_N \cap (\bigcup_{i \leq L} A_i)} h d\mu$$

$$< \frac{\epsilon}{2} + \int_{C_N \cap (\bigcup_{i \leq L} A_i)} (\sum_{n=1}^{N} f_n - \frac{\epsilon}{2\sum_{i \leq L} \mu(A_i)}) d\mu \leq \epsilon + \int_{C_N \cap (\bigcup_{i \leq L} A_i)} \sum_{n=1}^{N} f_n d\mu$$

$$= \epsilon + \sum_{n=1}^{N} \int_{C_N \cap (\bigcup_{i \leq L} A_i)} f_n d\mu \leq \epsilon + \sum_{n=1}^{N} \int_X f_n d\mu$$

$$\leq \epsilon + \sum_{n=1}^{\infty} \int_X f_n d\mu.$$

Letting $\epsilon \to 0$ we obtain

$$\int_X h d\mu \leq \sum_{n=1}^{\infty} \int_X f_n d\mu.$$

Since the integral of f is defined as the supremum of the integral of the simple functions $h \leq f$, this completes the proof of the claim. (Claim□)

□

Theorem 5.2 *Let (X, Σ) be a measure space. Let $\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ be a measure. Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of functions, each $f_n : X \to \mathbb{R}$ integrable. Suppose*

$$\sum_{n=1}^{n=\infty} \int |f_n| d\mu < \infty.$$

Then for μ-a.e. $x \in X$,

$$f(x) = \sum_{n=1}^{\infty} f_n(x)$$

is well defined, and moreover

$$\int f d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

Proof We have completed the special case of each $f_n \geq 0$ in 5.1 . Thus if let

$$g(x) = \sum_{n=1}^{\infty} |f_n(x)|$$

then we obtain that g is defined on a conull set, is integrable, and has $\int g d\mu = \sum_{n \in \mathbb{N}} \int |f_n| d\mu$. We then let $f(x) = \sum_{n \in \mathbb{N}} f_n(x)$ and have that f is well defined on all the points at which g is well defined. f will be integrable, because its absolute value is bounded by the integrable function g.

Fix $\epsilon > 0$. Appealing to 4.5 we find some measurable D with $\mu(D)$ finite and

$$\int_{X \setminus D} g d\mu < \frac{\epsilon}{5}.$$

At each N let D_N be the set

$$\{x \in D : \sum_{n=N}^{\infty} |f_n(x)| < \frac{\epsilon}{5\mu(D)}.$$

The D_N's are increasing and their union is conull in D. Thus we may find some N with $\int_{D \setminus D_N} g d\mu < \frac{\epsilon}{5}$.

Then at all $M \geq N$ we have

$$|\int_X f d\mu - \sum_{n=1}^{M} \int f_n d\mu| = |\int_X f d\mu - \int \sum_{n=1}^{M} f_n d\mu|$$

$$\leq \int_{D_N} |f - \sum_{n=1}^{M} f_n| d\mu + \int_{D \setminus D_N} |f| d\mu + \int_{D \setminus D_N} |\sum_{n=1}^{M} f_n| d\mu + \int_{X \setminus D} |f| d\mu + \int_{X \setminus D} |\sum_{n=1}^{M} f_n| d\mu$$

$$\leq \int_{D_N} \frac{\epsilon}{5\mu(D)} d\mu + \int_{D \setminus D_N} g d\mu + \int_{D \setminus D_N} g d\mu + \int_{X \setminus D} g d\mu + \int_{X \setminus D} g d\mu < \epsilon.$$

□

Theorem 5.3 *(*Monotone convergence theorem*) Let (X, Σ) be a measure space. Let $\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ be a measure. Let $(f_n)_{n \in \mathbb{N}}$ be sequence of functions, each $f_n : X \to \mathbb{R}$ integrable. Assume they are monotone, in the sense that either $f_n \leq f_{n+1}$ all n or $f_n \geq f_{n+1}$ all n. Suppose*

$$\int f_n d\mu$$

is bounded. Then there exists an integrable f with

$$f_n(x) \to f(x)$$

for μ-a.e. x and

$$\int |f_n - f| d\mu \to 0.$$

Proof Assume each $f_n \leq f_{n+1}$, since the other case is symmetrical. Note then that in fact the integrals

$$\int f_n d\mu$$

converge, since they form a bounded monotone sequence.

After possibly replacing each f_n by $f_n - f_1$ we may assume the functions are all positive. Let $g_1 = f_1$ and for $n > 1$ let $g_n = f_n - f_{n-1}$. Then

$$f_N = \sum_{n \leq N} g_n$$

and

$$\int f_N d\mu = \int \sum_{n \leq N} g_n = \sum_{n \leq N} \int g_n d\mu.$$

Now it follows from 5.2 that f is integrable and

$$\int f_n d\mu = \int \sum_{n=1}^{N} g_n d\mu = \sum_{n=1}^{N} \int g_n d\mu \to f.$$

Since $f \geq f_n$ at each n we have

$$\int |f - f_n| d\mu = \int f - f_n d\mu = \int f d\mu - \int f_n d\mu,$$

as required. □

Theorem 5.4 *(*Dominated convergence theorem*) Let (X, Σ) be a measure space. Let $\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ be a measure. Let $(f_n)_{n \in \mathbb{N}}$ be sequence of functions, each $f_n : X \to \mathbb{R}$ integrable. Suppose $g : X \to \mathbb{R}$ is an integrable function with $|f_n(x)| < g(x)$ all $x \in X$, $n \in \mathbb{N}$. Suppose $f : X \to \mathbb{R}$ is a function to which the f_n's converge pointwise – that is to say,*

$$f_n(x) \to f(x)$$

all $x \in X$.

Then f is integrable and

$$\int |f - f_n| d\mu \to 0.$$

Proof Fix $\epsilon > 0$. Apply 4.5 to find some C with $\mu(C) < \infty$ and $\int_{X \setminus C} g d\mu < \epsilon/6$. At each N we can let C_N be the set of $x \in C$ for which

$$\forall n \geq N (|f(x) - f_n(x)| < \frac{\epsilon}{3\mu(C)}).$$

The $_N$'s form an increasing set whose union is C, and thus we can find some large enough N with

$$\int_{C \setminus C_N} g d\mu < \frac{\epsilon}{6}.$$

Then

$$\int_X |f - f_N| d\mu < \int_{X \setminus C} |f - f_N| d\mu + \int_{C_N} |f - f_N| d\mu + \int_{C \setminus C_N} |f - f_N| d\mu$$

$$< \int_X 2g d\mu + \int_{C_N} \frac{\epsilon}{3\mu(C)} d\mu + \int_{C \setminus C_N} 2g d\mu$$

$$< \frac{2}{6\epsilon} + \frac{\mu(C_N)}{3\mu(C_N)} + \frac{2}{6\epsilon} = \epsilon.$$

□

Theorem 5.5 *(*Fatou's lemma*) Let (X, Σ) be a measure space. Let $\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ be a measure. Let $(f_n)_{n \in \mathbb{N}}$ be sequence of functions, each $f_n : X \to \mathbb{R}$ integrable, each $f_n \geq 0$. Suppose that*

$$\text{liminf}_{n \to \infty} \int f_n d\mu < \infty.$$

Then for a.e. $x \in X$

$$\text{liminf} f_n(x)$$

exists, and

$$\int \text{liminf}_{n \to \infty} f_n(x) d\mu \leq \text{liminf}_{n \to \infty} \int f_n d\mu.$$

Proof Let

$$g_n(x) = \inf\{f_m(x) : m \geq n\}.$$

Since $g_n(x)$ can be expressed as

$$\text{lim}_{m \to \infty} \text{min}\{f_{n+i}(x) : i \leq m\}$$

and $\min\{f_{n+i}(x) : i \leq m\} \geq \min\{f_{n+i}(x) : i \leq m+1\} \geq 0$ we can apply monotone convergence at 5.3 to get that each g_n is integrable with

$$\int g_n d\mu = \lim \int \min\{f_{n+i} : i \leq m\}.$$

$g_n \leq g_{n+1}$ and each $\int g_n d\mu$ is bounded by $\text{liminf}_{n \to \infty} \int f_n d\mu$, so we can apply monotone convergence once more to get that

$$\int \text{liminf}_{n \to \infty} f_n d\mu = \int \text{lim}_{n \to \infty} g_n d\mu = \text{lim}_{n \to \infty} \int g_n d\mu.$$

At each n and $k \geq n$ we have

$$g_n \leq f_k,$$

and hence

$$\int g_n d\mu \leq \text{inf}_{k \geq n} \int f_k d\mu,$$

and thus

$$\int \text{liminf}_{n \to \infty} d\mu = \int \text{lim}_{n \to \infty} g_n d\mu$$

$$= \text{lim}_{n \to \infty} \int g_n d\mu \leq \text{lim}_{n \to \infty} \text{inf}_{k \geq n} \int f_k d\mu = \text{liminf}_{n \to \infty} \int f_n d\mu.$$

□

Theorem 5.6 *(Egorov's theorem) Let (X, Σ) be a finite measure space. Let $(f_n)_n$ be a sequence of measurable functions which converge* pointwise *– that is to say there is a function f such that for all x*

$$f_n(x) \to f(x)$$

as $n \to \infty$. Then for any $\epsilon > 0$ there is $A \in \Sigma$ with $\mu(X \setminus A) < \epsilon$ and (f_n) converging uniformly *to f on A – that is to say*

$$\forall \delta > 0 \exists N \in \mathbb{N} \forall n > N \forall x \in A (|f_n(x) - f(x)| < \delta).$$

Proof At each $N \in \mathbb{N}$ and $\delta > 0$ we can let

$$B_{N,\delta} = \{x : \forall n, m > N(|f_n(x) - f_m(x)| < \delta\}.$$

For each δ and each N, $B_{N,\delta} \subset B_{N+1,\delta}$, each $B_{N,\delta}$ is measurable, and the union

$$\bigcup_N B_{N,\delta} = X.$$

Thus at each $k \geq 1$ we can find some N_k such that

$$\mu(X \setminus B_{N_k, \frac{1}{k}}) < 2^{-k}\epsilon.$$

Then for

$$B = \bigcap_k B_{N_k}$$

we have $\mu(X \setminus B) < \epsilon$ and for all $x \in B$ and all $n, m > N_k$

$$|f_n(x) - f_m(x)| < \frac{1}{k}$$

$$\therefore |f_n(x) - f(x)| \leq \frac{1}{k}.$$

$\square$

6 Radon-Nikodym and conditional expectation

One of the most important theorems in measure theory is Radon-Nikodym. It can be proved without a large amount of background and we may as well do so now.

Definition Let X be a set and $\Sigma \subset \mathcal{P}(X)$ a σ-algebra.

$$\mu : \Sigma \to \mathbb{R}$$

is said to be a *signed measure* if
(a) $\mu(\emptyset) = 0$;
(b) if $(A_n)_{n\in\mathbb{N}}$ is a sequence of disjoint sets in Σ, then

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

Note here we *are* assuming finiteness of the measure and in (b) above we are demanding convergence of the series. Here in fact (a) is redundant – following from (b) and μ only taking finite values.

Lemma 6.1 *If Σ is a σ-algebra on X and $\mu : \Sigma \to \mathbb{R}$ is a signed measure, then whenever $(B_n)_{n\in\mathbb{N}}$ is a sequence of sets in Σ*

$$\mu(\bigcap_{n\in\mathbb{N}} B_n) = \lim_{N\to\infty}\mu(\bigcap_{n\le N} B_n).$$

Proof First some cosmetic rearrangement. Let $C_n = \bigcap_{i\le n} B_i$. So at every n we have $C_n \supset C_{n+1}$, but the sequence has the same infinite intersection. Now consider the difference sets and define $D_n = C_n \setminus C_{n+1}$; the D_n's are now disjoint and if we let $B_\infty = \bigcap_{i\in\mathbb{N}} B_i$ represent the infinite intersection we have the equalities

$$C_n = B_\infty \cup D_n \cup D_{n+1} \cup D_{n+2}....$$

at every n. Thus

$$\mu(C_n) = \mu(B_\infty) + \sum_{m\ge n} \mu(D_m).$$

This in particular implies $\sum_{m=1}^{\infty} \mu(D_m)$ is convergent and

$$\sum_{m\ge n} \mu(D_m) \to 0$$

as $n \to \infty$, which is all we need to ensure $\mu(C_n) \to \mu(B_\infty)$. □

Theorem 6.2 *(Hahn Decomposition Theorem) Let Σ be a σ-algebra on X and $\mu : \Sigma \to \mathbb{R}$ a signed measure. Then there exists $A \in \Sigma$ such that for all $B \in \Sigma$, $B \subset A$*

$$\mu(B) \ge 0$$

and for all $C \in \Sigma$, $C \subset X \setminus A$

$$\mu(C) \le 0.$$

Proof Let δ be the supremum of the set $\{\mu(A) : A \in \Sigma\}$. Let $(B_n)_{n\in\mathbb{N}}$ be a sequence of sets in Σ with

$$\mu(B_n) \to \delta.$$

Then at each n let $\mathcal{A}_n$ be the algebra of sets generated by $(B_i)_{i\leq n}$.

Let's pause for a moment and observe some of the properties of these $\mathcal{A}_n$ algbras. First, each is finite, since it is generated by finitely many sets. Moreover $\mathcal{A}_n$ will have "atoms" of the form

$$\bigcap_{i\in S} B_i \cap \bigcap_{i\leq n, i\notin S} B_i :$$

each of these atoms contains no smaller non-empty set in $\mathcal{A}_n$ and every element of $\mathcal{A}_n$ is the finite union of such atoms. Finally note that B_n is an element of $\mathcal{A}_n$.

At each n, let C_n be element of $\mathcal{A}_n$ with maximum value under μ. Since $B_n \in \mathcal{A}_n$ we have

$$\mu(B_n) \leq \mu(C_n)$$

and $\mu(C_n) \to \delta$ and C_n consists of the finite union of atoms in $\mathcal{A}_n$ with positive measure.

Claim: At each n, $\mu(\bigcup_{i\geq n} C_i) \geq \mu(C_n)$.

Proof of Claim: Since at any $j \geq n$, $C_j \setminus \bigcup_{n\leq i<j} C_i$ equals the union of finitely many atoms with positive measure. (Claim□)

Now let

$$A = \bigcap_{n\in\mathbb{N}} \bigcup_{i\geq n} C_i.$$

Then by the last lemma $\mu(A) = \lim_{n\to\infty}\mu(\bigcup_{i\geq n} C_i) = \delta$.

Now note that is immediately implies δ is finite, since μ is finite and we attained the value δ with $\mu(A)$.

Since A has *attained* this maximum value δ we must have every $B \subset A$ in Σ with $\mu(B) \geq 0$ (for otherwise $\mu(A \setminus B)$ would be greater than $\mu(A)$). Similarly for any $C \in \Sigma$ disjoint to A we must have $\mu(C) \leq 0$. □

Theorem 6.3 *(Hahn-Jordan Decomposition) Let Σ be a σ-algebra on a set X. Let $\mu : \Sigma \to \mathbb{R}$ be a signed measure. Then we can find to measures $\mu^+, \mu^- : \Sigma \to \mathbb{R}^{\geq 0}$ with*

(i) $\mu = \mu^+ - \mu^-$;

(ii) μ^+, μ^- have disjoint support.

Proof The statement of the theorem should be quickly clarified. (i) states that for any set $C \in \Sigma$ we have $\mu(C) = \mu^+(C) - \mu^-(C)$. (ii) states that we can some $A \in \Sigma$ with $\mu^+(B) = 0$ when $B \subset X \setminus A$ and $\mu^-(B) = 0$ when $B \subset A$.

With this clarification in mind, the proof of the theorem is an immediate consequence of 6.2: Choose A for μ as there, and then let $\mu^+(B) = \mu(A \cap B)$ and $\mu^-(B) = -\mu(B \setminus A)$. □

Definition Given two measures $\mu, \nu : \Sigma \to \mathbb{R} \cup \{\infty\}$, we say that μ is *absolutely continuous with respect to ν*, written

$$\mu << \nu,$$

if whenever $B \in \Sigma$ has $\nu(B) = 0$ then $\mu(B) = 0$.

We have dealt with the concept of measurable functions in different contexts already. So there is no possibility of confusion, let us fix a convention for the entirely general context of σ-algebras.

Definition Let Σ be a σ-algebra on a set X. $f : X \to \mathbb{R}$ is *measurable with respect to* Σ if $f^{-1}[U]$ is in Σ for any open $U \subset \mathbb{R}$.

Exercise Show that $f : X \to \mathbb{R}$ is measurable with respect to Σ if for any $q \in \mathbb{Q}$ we have

$$f^{-1}[(-\infty, q)] \in \Sigma.$$

Exercise Show that if $f : X \to \mathbb{R}$ is measurable with respect to Σ then for any Borel $B \subset \mathbb{R}$ we have $f^{-1}[B] \in \Sigma$.

Theorem 6.4 *(Radon-Nikodym) Let Σ be a σ-algebra on a set X. Let $\mu, \nu : \Sigma \to \mathbb{R} \cup \{\infty\}$ be σ-finite measures with $\mu << \nu$. Then there is a measurable with respect to Σ function*

$$f : X \to \mathbb{R}$$

such that for any $C \in \Sigma$

$$\mu(C) = \int_C f(x) d\nu(x).$$

Proof We can assume μ, ν are finite measures, since otherwise we could partition the space X into a countable collection of elements in Σ on which both are finite and it would suffice to prove the theorem on each of these pieces.

At each $q \in \mathbb{Q}$, $q > 0$, let $\mu_q = \mu - q \cdot \nu$; that is to say, we define μ_q by

$$\mu_q(A) = \mu(A) - q\nu(A).$$

Each of these is a signed measure on (X, Σ). Applying 6.2 we can find $A_q \in \Sigma$ with $\mu_q(B) \geq 0$ all $B \subset A_q, B \in \Sigma$, $\mu_q(C) \leq 0$ all $C \subset X \setminus A_q, C \in \Sigma$.

Note that for $q_1 < q_2$ we have $A_{q_1} \setminus A_{q_2}$ null with respect to ν and hence after discarding some null sets we can assume

$$q_1 < q_2 \Rightarrow A_{q_2} \subset A_{q_1}.$$

By the assumption of $\mu << \nu$ we get $\bigcap_{q \in \mathbb{Q}} A_q$ null with respect to both these measures – since otherwise we could let $A_\infty = \bigcap_{q \in \mathbb{Q}} A_q$ and unwinding the definitions we would have $\mu(A_\infty) > q\nu(A_\infty)$ all $q \in \mathbb{Q}$, which would imply $\nu(A) = 0$; and then again after possibly discarding a null set we can assume

$$\bigcap_{q \in \mathbb{Q}} A_q = \emptyset;$$

so if we let

$$f(x) = \sup\{q : x \in A_q\}$$

we obtain a measurable with respect to Σ function $f : X \to \mathbb{R}^{\geq 0}$.

For $q_1 < q_2$ let $B_{q_1, q_2} = A_{q_1} \setminus A_{q_2}$.

Claim: For $B \subset B_{q_1, q_2}$ in Σ

$$|\int_B f(x) d\mu(x) - \mu(B)| \leq (q_2 - q_1)\nu(B).$$

Proof of Claim: We have $B \subset A_{q_1}$

$$\therefore \mu_{q_1}(B) \geq 0$$

$$\therefore \mu(B) \geq q_1 \nu(B),$$

and similarly B is disjoint to A_{q_2} and

$$\therefore \mu_{q_2}(B) \leq 0,$$

$$\therefore \mu(B) \leq q_2\nu(B).$$

In other words, we have the inequality

$$q_1\nu(B) \leq \mu(B) \leq q_2\nu(B).$$

Then since $f(x)$ ranges between q_1 and q_2 on B_{q_1,q_2} and hence B we obtain the parallel inequality

$$q_1\nu(B) \leq \int_B f(x)d\nu(x) \leq q_2\nu(B),$$

and hence

$$|\int_B f(x)d\mu(x) - \mu(B)| \leq (q_2 - q_1)\nu(B),$$

as required. (Claim□)

This last observation is all we need. Given any $C \subset X$ we can first fix $\epsilon > 0$ and let

$$C_\ell = C \cap B_{\ell\cdot\epsilon,(\ell+1)\cdot\epsilon}.$$

Again we are implicitly using $\mu << \nu$ to see that $C = \bigcup_{\ell\in\mathbb{N}} C_\ell$.

$$|\int_C f(x)d\nu(x) - \mu(C)| = |\sum_{\ell\in\mathbb{N}} \int_{C_\ell} f(x)d\nu(x) - \sum_{\ell\in\mathbb{N}} \mu(C_\ell)|$$

$$\leq \sum_{\ell\in\mathbb{N}} |\int_{C_\ell} f(x)d\nu(x) - \nu(C_\ell)|,$$

which by the above claim is bounded by

$$\sum_{\ell\in\mathbb{N}} \epsilon\nu(C_\ell) = \epsilon\nu(C).$$

Letting ϵ tend to zero we obtain $\int_C f(x)d\nu(x) = \nu(C)$. □

The function f we arrived at in the theorem above is not necessarily unique, but it is *almost* unique. I will leave at as an exercise for you to see that if f_0 is another function with

$$\mu(C) = \int_C f_0(x)d\nu(x)$$

on any $C \in \Sigma$ then f_0 agrees with f off a ν null set. This virtual uniqueness motivates a definition: We say that f as in the theorem is the *Radon-Nikodym derivative*, and is sometimes denoted by the slightly poetical notation

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

A crucial application of Radon-Nikodym is the existence of *conditional expectation.* At first sight the theorem may appear abstract to the point of being ethereal. A couple of motivating examples can give a sense of its true content.

Examples (i) Let X be a finite set – say $X = \{1, 2, 3, 4, 5, 6\}$. Let $f : X \to \mathbb{R}$. For instance, $f(n) = n^2$, just for example. Let B be a subset of X – say $B = \{2, 3, 4\}$. If someone tells you that they are thinking of a number chosen randomly from B, you would probably have an intuitive idea of the *expectation* of f on B: You would probably take the average value of f over B:

$$\frac{1}{3}(4 + 9 + 16) = 9\frac{2}{3}.$$

(ii) A little bit more abstract, let us take X to be the surface of the planet earth and $f(x)$ the average temperature of the location x. Using that information you would probably be able to go ahead and calculate the average temperature at a given latitude. So in this way we could discard, as it were, some of the information carried by f and obtain another function which records averages along the latitudes alone.

(iii) Alternatively, you may have a formula which can precisely compute the oxygen intake of a microbe based on its size and age. In the course of an experiment perhaps only the age is known, and then the best guess as to the oxygen intake would be your expectation given the partial information available.

Intuitively then it doesn't seem outrageous to give a best *guess* or *expectation* of a function on the basis of partial information. The following slick theorem justifies this rigorously. With Radon-Nikodym already available to us, the proof is very, very short – don't blink or you will miss it.

Theorem 6.5 *Let $\Sigma_0 \subset \Sigma_1$ be two σ-algebras on a set X. Let μ be a measure on (X, Σ_1). Assume that μ is σ-finite with respect to Σ_0 – we can partition X in to sets in Σ_0 on which μ is finite. Let*

$$f : X \to \mathbb{R}$$

be measurable with respect to Σ_1.

Then there is a function

$$g : X \to \mathbb{R}$$

which is measurable with respect to Σ_0 such that on any $B \in \Sigma_0$

$$\int_B g(x) d\mu(x) = \int_B f(x) d\mu(x).$$

Proof First of all, we can assume $f \geq 0$, since otherwise we write $f = f^+ - f^-$, $f^+ = \frac{1}{2}(|f| + f)$, $f^- = \frac{1}{2}(|f| - f)$ and apply the result to these two non-negative functions in turn.

Now let ν be defined on (X, Σ_0) by

$$\nu(B) = \int_B f(x) d\mu(x)$$

all $B \in \Sigma_0$. Let $\mu_0 = \mu|_{\Sigma_0}$, the restriction of μ to the sub σ-algebra Σ_0. Thus we have ν and μ_0 two measures on Σ_0. Clearly

$$\nu << \mu_0$$

since if $\mu(B) = 0$ then certainly $\int_B f(x) d\mu(x) = 0$.

Hence we can apply 6.4 and obtain $g : X \to \mathbb{R}$ which is measurable with respect to Σ_0 and has for all $B \in \Sigma_0$

$$\int_B f(x) d\mu(x) = \nu(B) = \int_B g(x) d\mu(x),$$

just as needed. $\square$

Definition For $f, g, \Sigma_0, \Sigma_1, X$ as in the statement of the last theorem, we say that g is *the conditional expectation of f with respect to* Σ_0 and write

$$g = E(f|\Sigma_0).$$

Strictly speaking there is the same grammatical flaw in this terminology which we saw in the use of the term *the* Radon-Nikodym derivative. The conditional expectation is only defined up to null sets, but since this is good enough for our purpose we indulge the definite article.

Think of $E(f|\Sigma_0)$ this way: This is the function whose value at a point $x \in X$ only depends on which elements of Σ_0 the point lies inside; it is as if we are forbidden to access any information about x which uses sets in Σ_1 but not Σ_0.

In passing we mention that 6.3 gives a transparent definition of integration against signed measures.

Definition Let Σ be a σ-algebra on a set X and let μ be a signed measure on Σ. Let $\mu^+, \mu^- : \Sigma \to \mathbb{R}$ be measures on Σ with

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

and μ^+, μ^- having disjoint supports. Then for any Σ-measurable $f : X \to \mathbb{R}$ we let

$$\int f d\mu = \int f d\mu^+ - \int f d\mu^-.$$

7 Standard Borel spaces

The theory of measurable spaces can be developed in various levels of generality. I generally take the view that most of the natural spaces in this context are either *Polish* spaces or *standard Borel* spaces.

Definition A topological space is *Polish* if it is separable and admits a compatible complete metric. We then define the *Borel sets* in the space to be those appearing in the smallest σ-algebra containing the open sets.

Examples (i) Any compact metric space forms a Polish space. For instance if we take

$$2^{\mathbb{N}} = \prod_{\mathbb{N}} \{0, 1\},$$

the countable product of the two element discrete space $\{0, 1\}$, then we have a Polish space. (For the metric, given $\vec{x} = (x_0, x_1, ...), \vec{y} = (y_0, y_1, ...)$, take $d(\vec{x}, \vec{y})$ to be 2^{-n}, where n is least with $x_n \neq y_n$.)

(ii) $\mathbb{R}$ and $\mathbb{C}$ are Polish spaces, as are all the $\mathbb{R}^N$'s and $\mathbb{C}^N$'s.

(iii) Any closed subset of a Polish space is Polish.

(iv) Let $C([0, 1])$ be the collection of continuous functions from the unit interval to $\mathbb{R}$. Given $f, g \in C([0, 1])$ let $d(f, g)$ be

$$\sup_{z \in [0,1]} |f(z) - g(z)|.$$

As some of you may be aware, this metric can be shown to be complete and separable, and hence the induced topology is Polish. (More generally, any separable Banach space is Polish in the topology induced by the norm.)

Exercise (i) The Borel subsets of a Polish space can be characterised as the smallest collection containing the open sets, the closed sets, and closed under the operations of countable union and countable intersection.

(ii) The Borel sets may also be characterised as the smallest collection containing the open sets, closed under complements, and closed under countable intersections.

Note here we don't care about the specific metric: Only that a complete compatible metric exists. We have abstracted away the metric, and only ask that the remaining topology could have been presented as arising from a suitable metric.

There is a bit of knack to showing sets are Borel. The next exercise is typical of the kind of reasoning we use – breaking down a seemingly complicated set into smaller constituents of its definitions.

Exercise Let $X = 2^{\mathbb{N}}$, the collection of all functions from $\mathbb{N}$ to $\{0, 1\}$ in the product topology.

(i) Show that for each N and $r \in \mathbb{R}$, the collection $A_{N,r}$ of $f \in X$ with

$$\frac{1}{N} |\{n < N : f(n) = 1\}| > r$$

is Borel.

(ii) Similarly, for each N and $r \in \mathbb{R}$, the collection $B_{N,r}$ of $f \in X$ with

$$\frac{1}{N} |\{n < N : f(n) = 1\}| < r$$

is Borel.

(iii) Show that the set of $f \in X$ such that

$$\text{liminf}_{N \to \infty} \frac{1}{N} |\{n < N : f(n) = 1\}| \geq \frac{1}{2}$$

is Borel. (Hint: $\bigcap_{q<\frac{1}{2},q\in\mathbb{Q}}\bigcup_{M\in\mathbb{N}}\bigcap_{N\geq M}A_{N,q}$.)

(iv) Similarly for limsup $\leq \frac{1}{2}$.

(v) And thus the set of $f \in X$ with

$$\lim_{N\to\infty}\frac{1}{N}|\{n<N : f(n)=1\}| = \frac{1}{2}$$

is Borel.

Very frequently we are only concerned with a Polish space's Borel structure. This prompts one further round of abstraction.

Definition A set X equipped with a σ-algebra Σ is said to be a *standard Borel space* if there is some choice of a Polish topology on X which gives rise to Σ as the corresponding collection of Borel sets.

At the end of these short definitions there is a remarkable fact whose proof is too involved to present here.

Theorem 7.1 *Any two uncountable standard Borel spaces are Borel isomorphic.*

That is to say, if $(X_1,\Sigma_1),(X_2,\Sigma_2)$ are uncountable Borel spaces, then there is a bijection

$$\pi : X_1 \to X_2$$

such that for all $A \subset X_2$

$$A \in \Sigma_2 \Leftrightarrow \pi^{-1}[A] \in \Sigma_1.$$

A proof of this theorem can be found at 15.6 [6]. A key part of the proof is showing any uncountable Polish space contains a homeomorphic copy of Cantor space.

Definition If X is a standard Borel space and $\Sigma \subset \mathcal{P}(X)$ the σ-algebra of Borel sets, then a *Borel measure* on X is a measure in the earlier sense

$$\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}.$$

Again we say that $M \subset X$ is *measurable* if there are Borel sets A, B with

$$A \subset M \subset B,$$

$$\mu(B \setminus A) = 0.$$

We then let $\mu(M) = \mu(A)$.

Definition A measure μ on X is said to have a point $a \in X$ as an *atom* if $\mu(\{a\}) > 0$. μ is said to be a *probability measure* if $\mu(X) = 1$. A standard Borel space equipped with a Borel probability measure is called a *standard Borel probability space.*

And again we have a remarkable fact:

Theorem 7.2 *Any two atomless standard Borel probability spaces are isomorphic.*

In other words, if $(X_1, \Sigma_1, \mu_1), (X_2, \Sigma_2, \mu_2)$ are atomless standard Borel probability spaces, then there is a bijection

$$\pi : X_1 \to X_2$$

sending inducing an isomorphism of $(X_1, \Sigma_1) \cong (X_2, \Sigma_2)$ with the *additional property* that for $A \in \Sigma_2$

$$\mu_2(A) = \mu_1(\pi^{-1}[A]).$$

In particular, any atomless standard Borel probability space is isomorphic to the unit interval equipped with the (restriction of) Lebesgue measure. See 17.41 [6].

In this course I will largely concentrate on probability spaces. In the literature, almost all work on infinite measure spaces is under the assumption of the space being σ-finite – and in this case we can partition the space into countably many pieces of measure one. In the case of finite spaces with measure other than one, we can rescale the measure by a constant to get the total mass back to one. Thus the loss of generality in working with probability spaces is very minor.

Exercise (i) Show that if (X, Σ, μ) is a standard Borel probability space and $(B_n)_{n\in\mathbb{N}}$ is a sequence of sets in Σ, then

(a) $\mu(\bigcup_{n\in\mathbb{N}} B_n) = \lim_{N\to\infty}\mu(\bigcup_{n\leq N} B_n)$;

(b) $\mu(\bigcap_{n\in\mathbb{N}} B_n) = \lim_{N\to\infty}\mu(\bigcap_{n\leq N} B_n)$.

(ii) Show that (b) above might fail if we simply assume μ to be a σ-finite measure on a standard Borel space (X, Σ).

Definition A function $f : X \to Y$ between two Polish spaces is said to be *Borel* if $f^{-1}[U]$ is Borel for any open $U \subset Y$.

Lemma 7.3 *If $f : X \to Y$ is Borel, then for any Borel $B \subset Y$ we have $f^{-1}[B]$ Borel.*

Proof Let Σ be the collection of subsets of Y for which the pullback along f is Borel. By assumption this contains the open sets. It is a σ-algebra, since

$$f^{-1}[Y \setminus B] = X \setminus f^{-1}[B]$$

and

$$f^{-1}[\bigcap B_n] = \bigcap f^{-1}[B_n].$$

Thus it includes the Borel sets. □

With this lemma in our tool kit we can go forward and define the concept of Borel function when there is no topology in sight.

Definition Let X and Y be standard Borel spaces. $f : X \to Y$ is a *Borel function* if $f^{-1}[B]$ is Borel for any Borel $B \subset Y$.

Exercise (i) Show that if X and Y are Polish spaces, then $X \times Y$ is Polish in the product topology.

(ii) For X and Y as above, and $f : X \to Y$ a Borel function, show that f is Borel as a subset of $X \times Y$.

Lemma 7.4 *$f : X \to \mathbb{R}$ is measurable if $f^{-1}[O]$ is measurable for each open $O \subset \mathbb{R}$.*

Proof Let Σ be the collection of all $A \subset \mathbb{R}$ with $f^{-1}[A]$ measurable. Σ includes the open sets by assumption. Since the measurable subsets of X form a σ-algebra and

$$f^{-1}[\mathbb{R} \setminus A] = X \setminus f^{-1}[A],$$
$$f^{-1}[\bigcup A_i] = \bigcup f^{-1}[A_i],$$
$$f^{-1}[\bigcap A_i] = \bigcap f^{-1}[A_i],$$

we have that Σ includes the Borel sets. $\square$

Exercise Show that any measurable function on a standard Borel probability space agrees with a Borel function on some conull Borel set.

This concept of measure is apparently ethereal. It involves considering the collection of all Borel subsets and considering the behavior of the measure with respect to arbitrary sequences of Borel sets. Later in the course we will discuss a form of the *Riesz representation theorem* which enables us to give a concrete description of the collection of Borel probability measures on a compact metric space; this representation theorem will in particular enable us to view the collection of probability measures as forming a well behaving topological space in its own right.

8 Borel and measurable sets and functions

It is only a slight exaggeration to describe standard Borel probability spaces as the basic object of study for this course. Recall that this consists of a set X equipped with a σ-algebra Σ and a function

$$\mu : \Sigma \to [0, 1]$$

such that there is a Polish topology on X giving rise to Σ as the Borel sets and μ is a measure with $\mu(X) = 1$.

Of course people can and do study measures on more general kinds of spaces, but for practical purposes standard Borel spaces are likely to include all the examples you will ever encounter. The situation with the measure having mass one – $\mu(X) = 1$ – is a bit more subtle. It is natural to consider σ-finite measures – such as $\mathbb{R}$ equipped with Lebesgue measure. But even here one can partition the space into countably many pieces each having measure one. In the case that $\mu(X)$ is a finite number other than 1, the situation is sufficiently similar to a probability space for us to be unconcerned.

This section will discuss finer results on measurable functions on standard Borel spaces. We could spend a lot more time here, and a lot of the results are on the edge of my own field of descriptive set theory. Instead I will present a couple of tricks which occur over and over in certain branches of analysis. Frequently one needs to know when a function or set is measurable. Certainly Borel sets are measurable; we will also see that the projections of Borel sets are measurable as well.

Lemma 8.1 *Let X be a Polish space, Σ its σ-algebra of Borel sets, and μ a Borel probability measure on X. Then for $A \in \Sigma$ we have*

$$\begin{aligned} \mu(A) &= \sup(\{\mu(F) : F \subset A, F \text{ closed}\}) \\ &= \inf(\{\mu(O) : O \supset A, O \text{ open}\}). \end{aligned}$$

Proof Let d be a compatible complete metric on X. Let Σ_0 be the collection consisting of all Borel sets A satisfying the conditions above. Note that for O open and $n \in \mathbb{N}$ we can let

$$F_n = \{x \in O : d(x, X \setminus O) \geq \frac{1}{n}\},$$

where $d(x, X \setminus O) = \inf\{d(x, y) : y \in X \setminus O\}$. Then

$$O = \bigcup_n F_n,$$

and hence $\mu(O) = \lim \mu(F_n)$, and thus $O \in \Sigma_0$.

Inspecting the definition and using $\mu(X) = 1$ we have that Σ_0 is closed under complements.

Finally suppose $(A_n)_{n\in\mathbb{N}}$ is a sequence of sets in Σ_0. Fix $\epsilon > 0$. For each n we can find closed $F_n \subset A_n$ with $\mu(A_n \setminus F_n) < \frac{\epsilon}{2^{n+1}}$. Then we can go to some large K with $\mu((\bigcup_{n\in\mathbb{N}} A_n) \setminus (A_1 \cup A_2 \cup ...A_K)) < \frac{\epsilon}{2}$. It then follows that $\mu((\bigcup_{n\in\mathbb{N}} A_n) \setminus F_1 \cup F_2...F_k) < \epsilon$.

Similarly if we choose open $O_n \supset A_n$ open with $\mu(O_n \setminus A_n) < \frac{\epsilon}{2^n}$, then $\mu((\bigcup_{n\in\mathbb{N}} O_n) \setminus (\bigcup_{n\in\mathbb{N}} A_n)) < \epsilon$..
□

Lemma 8.2 *Let X be a Polish space, Σ its σ-algebra of Borel sets, and μ a Borel probability measure on X. Then for $A \in \Sigma$ we have*

$$\mu(A) = \sup(\{\mu(K) : K \subset A, K \text{ compact}\}).$$

Proof Fix $\epsilon > 0$. By the last lemma we may assume A is closed. Now fix $(x_i)_{i\in\mathbb{N}}$ dense in A and at each i let B_i^1 be $\{y \in A : d(x_i, y) \leq 1\}$. Thus we have covered A by countably many balls of radius 1. We may find some N_1 such that $\mu(A \setminus \bigcup_{i<N_1} B_i^1) < \frac{\epsilon}{2}$. Repeating we may let $B_i^2 = \{y \in A : d(x_i, y) \leq 1/2\}$ and find N_2 such that $\mu(A \setminus \bigcup_{i<N_2} B_i^2) < \frac{\epsilon}{4}$, and that at each ℓ set $B_i^\ell = \{y \in A : d(x_i, y) \leq 2^{-\ell}\}$ and find N_ℓ such that $\mu(A \setminus \bigcup_{i<N_\ell} B_i^\ell) < \frac{\epsilon}{2^\ell}$. If we take

$$K = \bigcap_\ell \bigcup_{i<N_\ell} \overline{B_i^\ell}$$

then K is closed and $2^{-\ell}$-bounded for each ℓ, hence compact. $\square$

Theorem 8.3 *(Lusin) Let X be a Polish space and μ a Borel probability measure on X. If*

$$f : X \to Y$$

is a Borel function into a Polish Y, then for any $\epsilon > 0$ there is a compact $K \subset X$ with $f|_K$ continuous and $\mu(K) > 1 - \epsilon$.

Proof Let $\{U_\ell : \ell \in \mathbb{N}\}$ be a countable basis for Y. At each ℓ apply 8.1 to find open $O_\ell \supset f^{-1}[U_\ell]$ with

$$\mu(O_\ell \setminus f^{-1}[U_\ell]) < \frac{\epsilon}{2^\ell}.$$

Then it is immediate that f is continuous on

$$X \setminus (\bigcup O_\ell \setminus f^{-1}[U_\ell]).$$

The measure of this set is greater than $1 - \epsilon$ and so we can finish by 8.2. $\square$

Definition For (A, d) a metric space and $B \subset A$, the *diameter* of B, $d(B)$, is the sup $d(a, a')$ as a, a' range over B.

Notation For s a sequence of length ℓ, and a an element, $s^\frown a$ is the sequence of length $\ell + 1$ extending s with final term a.

Lemma 8.4 *Let X, Y be Polish and*

$$f : X \to Y$$

continuous. Then $f[X]$ is measurable (with respect to any Borel probability measure μ on Y).

Proof Fix d_X and d_Y complete, compatible metrics on X and Y.

Claim: If $C \subset X$ closed and $\epsilon > 0$, then we can find $(C_n)_{n\in\mathbb{N}}$ closed subsets of C such that

$$C \subset \bigcup C_n$$

and at each n,

$$d_X(C_n) < \epsilon,$$
$$d_Y(f[C_n]) < \epsilon.$$

Proof of Claim: Fix a countable basis for X. Around each $x \in C$ we can find a *basic* open U of diameter less than ϵ with $f[U]$ of diameter less than ϵ. We then let $(C_n)_{n\in\mathbb{N}}$ enumerate the sets of the form

$$\overline{C \cap U}$$

as U ranges over such basic open sets. (□Claim)

Iterating the above lemma we can find an array of closed sets

$$(C_s)_{s\in\mathbb{N}^{<\infty}},$$

indexed by the finite sequences of natural numbers, such that

(a) $C_\emptyset = X$;
(b) each $\bigcup_{n\in\mathbb{N}} C_{s^\frown n} = C_s$;
(c) if $s \in \mathbb{N}^n$ (i.e. a sequence of length n), then

$$d_X(C_s) < 2^{-n},$$

$$d_Y(f[C_s]) < 2^{-n}.$$

At each s choose Borel $B_s \supset f[C_s]$ Borel with

$$B_s \subset \overline{f[C_s]}$$

and $\mu(B_s)$ as small as possible.

Claim $B_s \setminus \bigcup_{n\in\mathbb{N}} B_{s^\frown n}$ is always null.

Proof of Claim: Since

$$f[C_s] = \bigcup_{n\in\mathbb{N}} f[C_{s^\frown n}] \subset \bigcup_{n\in\mathbb{N}} B_{s^\frown n}$$

and we chose B_s to have minimal measure. (□Claim)

We then let $A = \bigcup_{s\in\mathbb{N}^{<\infty}}(B_s \setminus \bigcup_n B_{s^\frown n})$. A is the countable union of null sets, and hence null.
$B_\emptyset \supset f[C_\emptyset] = f[X]$ so it suffices to show $B_\emptyset \setminus A \subset f[C_\emptyset]$. If $y \in B_\emptyset \setminus A$, then we can choose

$$s_0 \subset s_1 \subset s_2 ...$$

such that each s_ℓ is of length ℓ and $y \in B_{s_\ell}$.

Then we may find $y_\ell \in f[C_{s_\ell}]$ since $B_{s_\ell} \subset \overline{f[C_{s_\ell}]}$ is non-empty. Then fix corresponding $x_\ell \in C_{s_\ell}$ with $f(x_\ell) = y_\ell$.

The diameters of the $(f[C_{s_\ell}])_\ell$ sets are approaching zero, so we have $y_\ell \to y$. Since $d_X(C_{s_\ell}) < 2^{-\ell}$ we have $(x_\ell)_\ell$ Cauchy and hence there is some x with $x_\ell \to x$. By continuity, $f(x) = y$. □

This lemma can in fact be proved even in the case when f is simply a Borel function. There are various ways of approaching the proof of the more general result, but I will proceed by showing any Borel function can be made continuous by an appropriate strengthening of the topology.

Lemma 8.5 *Let (X, d) be a complete metric space. Then there is a compatible complete metric bounded by 1.*

Proof Let $d^*(x, y) = \min(1, d(x, y))$. □

Theorem 8.6 *Let X be a Polish space and $B \subset X$ Borel. Then there is a stronger Polish topology on X under which B becomes clopen – that is to say, both open and closed.*

Here when I say one topology is *stronger* than another I mean it has more open sets.

Proof The usual pattern. We want to show that the collection Σ of such sets includes the open sets, is closed under complements and countable intersections.

First to see that it includes the open sets, let $O \subset X$ be open and d a compatible complete metric on X bounded by 1. For $x, y \in X \setminus O$ let $d'(x,y) = d(x,y)$. For $x \in O, y \in X \setminus O$, set $d'(x,y) = 2$. And finally for $x, y \in O$ let

$$d'(x,y) = \min\{1, d(x,y) + |\frac{1}{d(x, X \setminus O)} - \frac{1}{d(y, X \setminus O)}|\}.$$

It is easily seen that if $(x_n)_n$ is a d'-Cauchy sequence included in O, then it is d-Cauchy with limit inside O.

It is immediate from the structure of the definitions that Σ is closed under complements.

Finally, let $(B_n)_n$ be a sequence of sets in Σ and at each n let d_n be a complete metric which gives rise to a stronger Polish topology in which B_n is clopen. We may assume d_n is bounded by 1. We can then let

$$d^*(x,y) = \sum 2^{-n} d_n(x,y).$$

It is routine to verify this is a complete metric and that the resulting topology is separable. The topology generated by d^* is at least as fine as each of d_n's, so each B_n becomes clopen. Thus $\bigcap B_n$ is closed in the new Polish topology. Going back to the second paragraph of the proof, we can find a stronger Polish topology in which $\bigcap B_n$ becomes clopen. □

Corollary 8.7 *Let $f : X \to Y$ be a Borel function between Polish spaces. Then there is a stronger Polish topology on X under which f becomes continuous.*

Proof Let $\{U_n : n \in \mathbb{N}\}$ be a countable basis for the topology on Y. At each n let d_n be a complete metric on X given rise to a stronger Polish topology in which $f^{-1}[U_n]$ is clopen. We may assume each d_n is bounded by 1, and so the topology generated by the metric

$$d^*(x,y) = \sum 2^{-n} d_n(x,y)$$

is as required. □

Corollary 8.8 *If $f : X \to Y$ is Borel, $B \subset X$ Borel, then $f[B]$ is measurable in Y with respect to any Borel probability measure.*

Proof By 8.4 and 8.7. □

The corollary 8.8 or the next result below is sometimes called *Jankov von Neumann.*

There is a little bit more we can extract from the proof of 8.8 and 8.4. This extra piece turns out to be important in certain contexts. Roughly speaking it states that we may find a measurable *selector* for Borel functions – a right inverse if you will.

Theorem 8.9 *Let X and Y be standard Borel spaces, μ a standard Borel probability measure on Y, $f : X \to Y$ Borel. Then there is a measurable function $\rho : f[X] \to X$ such that*

$$f(\rho(y)) = y$$

all $y \in f[X]$.

Proof Fix compatible Polish topologies on X and Y. Following 8.7 we may assume f is continuous. Going through the proof of 8.4, it suffices to define ρ on

$$B_\emptyset \setminus \bigcup_s (B_s \setminus \bigcup_n B_{s^\frown n}).$$

But now for y in this set we can successively define n_0^y to be least n with

$$y \in B_{\langle n \rangle},$$

then n_1^y least n with

$$y \in B_{\langle n_0^y, n \rangle},$$

and more generally $n_{\ell+1}^y$ to be least such

$$y \in B_{\langle n_0^y, n_1^y, \ldots, n_\ell^y, n \rangle}.$$

One verifies that for each $s \in \mathbb{N}^{<\infty}$ and $\ell \in \mathbb{N}$ the set

$$\{y \in f[X] : \langle n_0^y, n_1^y, \ldots n_\ell^y \rangle = s\}$$

is measurable. We then let $\rho(y)$ be the unique x in

$$\bigcap_\ell C_{\langle n_0^y, n_1^y, \ldots, n_\ell^y \rangle}.$$

The function ρ is measurable since for any open $U \subset X$ we have $\rho(y) \in U$ if and only if there is some ℓ with

$$C_{\langle n_0^y, n_1^y, \ldots, n_\ell^y \rangle} \subset U.$$

□

Warning: There is a notorious paper from early last century where Lebesgue claimed the Borel image of a Borel set is always Borel. This is false. The counterexamples are not obvious, but they do exist.

What is true however is a rather subtle result when all the sections are sufficiently small. Recall that a function is *countable to one* if the preimage of every point in the range is finite or countably infinite.

Theorem 8.10 *(Lusin Novikov) Let X and Y be standard Borel spaces. Let $f : X \to Y$ be a countable to one Borel function. Then:*

(I) $f[X]$ is Borel;

(II) there is a countable collection of Borel functions $\{g_n : n \in \mathbb{N}\}$ from $f[X]$ to X, such that for all $y \in f[X]$,

$$\{g_n(y) : n \in \mathbb{N}\} = \{x \in X : f(x) = y\}.$$

The proof of this theorem, which goes far beyond the scope of our course, can be found at 18.10 [6].

9 Summary of some important terminology

This subject is riddled with many similarly sounding definitions. In some cases similar sounding phrases can mean the same thing, slightly different things, or even completely different things. The terminology used by mathematicians in the area is often somewhat illogical or contradictory, and it has come about for largely accidental reasons. However it is also a fact of life.

Here is a brief summary of some of the concepts we have considered so far.

Definition A set X equipped with a σ-algebra Σ is said to be a *measure space*. In this case some people might say that a subset of X is *measurable* if it is in Σ – but I have tried to avoid that terminology due to the potential confusion with the the radically different notion of μ-measurable, defined below.

A set X equipped with a σ-algebra Σ is said to be a *standard Borel space* if there is some choice of a Polish topology on X which gives rise to Σ as its collection of Borel sets – in other words, Σ is the smallest σ-algebra containing the open sets of the Polish topology on X. Here one customarily refers to the elements of Σ as the *Borel sets*. Be warned that people will often refer to X as a standard Borel space without making explicit reference to Σ; this is similar to identifying a topological space with its underlying set or identifying a group with its set of elements – not quite literally and fastidiously correct, but natural once you are comfortable with the concepts.

Given a measure space (X, Σ) and a measure μ on Σ, we say that $A \subset X$ is μ-measurable if there are $B_1, B_2 \in \Sigma$ with $B_1 \subset A \subset B_2$ and $\mu(A_2 \setminus A_1) = 0$. Unfortunately in this situation if the measure μ is made clear by context, some people may simply refer to the μ-measurable sets as "measurable", but because of the obvious potential for confusion, I have tried to avoid doing so.

In the context of Lebesgue measure m_n on $\mathbb{R}^n$, we say that $A \subset \mathbb{R}^n$ is *Lebesgue measurable* if it is m_n-measurable in the sense above.

Definition Let $(X_1, \Sigma_1), (X_2, \Sigma_2)$ be measure spaces. A function

$$f : X_1 \to X_2$$

is *measurable* if for any $A \in \Sigma_2$ we have $f^{-1}[A] \in \Sigma_1$. If there is any uncertainty about which σ-algebra on X_1 we have in mind, one would refer to such an f as a Σ_1-*measurable* function.

Given $(X_1, \Sigma_1), (X_2, \Sigma_2)$ standard Borel,

$$f : X_1 \to X_2$$

is *Borel* if for any Borel subset A of X_2 we have $f^{-1}[A]$ Borel as a subset of X_1. Thus a Borel function is a measurable function where the measure spaces in question are standard Borel. Be warned that some authors use *Borel measurable function* to refer to what I am calling a "Borel function".

Given $(X_1, \Sigma_1), (X_2, \Sigma_2)$ measure spaces and μ a measure on Σ_1, we say that a function

$$f : X_1 \to X_2$$

is μ-*measurable* if for any $A \in \Sigma_2$ we have that $f^{-1}[A]$ is μ-measurable. Alas, in the strange, arbitrary world in which we live in, some people will refer to such a function f as simply being "measurable" if context has indicated μ – which clearly conflicts with the definition of measurable function between measure spaces I have given above.

There are various equivalent conditions on a function being measurable, Borel, or μ-measurable.

Lemma 9.1 *Let $(X_1, \Sigma_1), (X_2, \Sigma_2)$ be standard Borel spaces. Let τ be a Polish topology on X_1 which gives rise to Σ_1 as its collection of Borel sets. Let $\mathcal{B}$ be a basis for the topology τ. Then the following are equivalent:*

1. *f is Borel;*
2. *$f^{-1}[V] \in \Sigma_1$ for any $V \in \tau$;*
3. *$f^{-1}[V] \in \Sigma_1$ for any $V \in \mathcal{B}$.*

Lemma 9.2 *Let $(X_1, \Sigma_1), (X_2, \Sigma_2)$ be standard Borel spaces and let μ be a measure on Σ_1. Let τ be a Polish topology on X_1 which gives rise to Σ_1 as its collection of Borel sets. Let $\mathcal{B}$ be a basis for the topology τ. Then the following are equivalent:*

1. *f is μ-measurable;*
2. *$f^{-1}[V]$ is μ-measurable for any $V \in \tau$;*
3. *$f^{-1}[V]$ is μ-measurable for any $V \in \mathcal{B}$.*

10 Review of Banach space theory

It goes too far beyond the scope of this course to enter into any proofs in this section. I will, however, recall the basic material we will need for the discussion of the Riesz representation theorem and Stone-Weierstrass.

Banach spaces can be taken over either the field $\mathbb{R}$ or the field $\mathbb{C}$. The former is conceptually simpler to work with, but in some cases it is important to be able to find roots for polynomials, and that requires $\mathbb{C}$. I will use $\mathbb{K}$ to denote either $\mathbb{R}$ or $\mathbb{C}$, depending on the circumstances.

Definition Let $\mathbb{K}$ be either $\mathbb{R}$ or $\mathbb{C}$. A vector space $\mathbb{B}$ over $\mathbb{K}$ is said to be a *Banach space* if it is equipped with a function

$$\|\cdot\| : \mathbb{B} \to \mathbb{R}^{\geq 0}$$

such that:

1. $\|v\| = 0$ if and only if $v = 0$;
2. $\|v + w\| \leq \|v\| + \|w\|$;
3. $\|cv\| = |c|\|v\|$ for $c \in \mathbb{K}$, $v \in \mathbb{B}$;
4. if we define $d(v, w) = \|v - w\|$, then $d(\cdot, \cdot)$ is a complete metric on $\mathbb{B}$.

Examples

1. An example which has been lurking in the background is $L^1(X, \mu)$, where (X, Σ) is a measure space and $\mu : \Sigma \to \mathbb{R}^{\geq 0} \cup \{\infty\}$ is a measure. This consists of all integrable functions $f : X \to \mathbb{K}$ with $\|f\| = \int_X |f(x)| d\mu(x)$. In order to make this pass item 1 from the definition of a Banach space we identify any two functions which agree a.e. There is a non-trivial argument to show completeness of the norm – see for instance [7] or [10].

2. A generalization of this is to the L^p spaces for $p \geq 1$. Here $L^p(X, \mu)$ is the collection of functions for which $x \mapsto |f(x)|^p$ is integrable, with norm $\|f\| = (\int_X |f(x)|^p d\mu(x))^{\frac{1}{p}}$.

   Here the example of $L^2(X, \mu)$ is especially important, because it forms a *Hilbert space*. We have an inner product

   $$\langle f, g \rangle = \int_X f(x)\overline{g(x)} d\mu.$$

   (Here $\overline{g(x)}$ is the complex conjugate of $g(x)$. If $g(x) = a + bi, a, b \in \mathbb{R}$, then its complex conjugate is $a - bi$.)

3. A slight variation on the above example is $L^\infty(X, \mu)$ – where we can think of this as corresponding to $p = \infty$. This is the collection of measurable functions $f : X \to \mathbb{K}$ with f *essentially bounded* – which is to say, for some $c \in \mathbb{R}^{>0}$ we have $|f(x)| < c$ for μ-a.e. x.

4. There is a discrete variation of the above spaces. Given S some set, we let $\ell^p(S)$ be the set of functions $f : S \to \mathbb{K}$ such that

   $$\sum_{a \in S} |f(a)|^p < \infty.$$

   Note that in particular any $f \in \ell^p(S)$ will be identically zero off of a countable set.

5. For K a compact metric space, $C(K)$ equals the collection of continuous functions from K to $\mathbb{K}$. A continuous function on a compact space is bounded, hence we obtain a well defined non-negative quantity if we let $\|f\| = \sup\{|f(x)| : x \in K\}$. The argument that the resulting metric is complete on $C(K)$ amounts to a classical theorem from real analysis that a uniformly convergent sequence of uniformly continuous functions converges to a continuous function. (We will return to this example in greater detail in the section on the Riesz representation theorem).

Definition For $\mathbb{B}$ a Banach space, $\mathbb{B}^*$, the *dual of* $\mathbb{B}$, consists of all linear functions $\phi : \mathbb{B} \to \mathbb{K}$ which are *bounded* in the sense that

$$\{\frac{|\phi(v)|}{\|v\|} : v \in \mathbb{B}, v \neq 0\}$$

is bounded. We then let $\|\phi\| = \sup\{\frac{|\phi(v)|}{\|v\|} : v \in \mathbb{B}, v \neq 0\}$. We equip $\mathbb{B}^*$ with linear operations given by pointwise addition and scalar multiplication:

$$\varphi_1 + \varphi_2 : v \mapsto \varphi_1(v) + \varphi_2(v),$$

$$c\varphi : v \mapsto c \cdot \varphi(v).$$

In these operations and the indicated norm, $\mathbb{B}^*$ itself becomes a Banach space.

Examples In many specific situations there are very concrete ways of viewing the dual space.

1. If $\mathbb{B} = L^2(X, \mu)$ from above, then $\varphi : L^2(X, \mu) \to \mathbb{R}$ will be in the dual if and only if there is some $f \in L^2(X, \mu)$ such that
$$\varphi(g) = \int_X f(x)g(x)d\mu(x).$$

2. The dual for $L^1(X, \mu)$ turns out to be $L^\infty(X, \mu)$, in the sense that
$$\varphi : L^1(X, \mu) \to \mathbb{K}$$
will be $(L^1(X, \mu))^*$ if and only if there is some $f \in L^\infty(X, \mu)$ such that
$$\varphi(g) = \int_X f(x)g(x)d\mu$$
for all $g \in L^1(X, \mu)$.

3. In 11§ we will see that the dual of $C(K)$ can be identified with the signed or complex valued finite measures on K.

Definition In general there are always two distinct topologies we can define on a Banach space, and in the specific case that it is a dual space there is a third. For $\mathbb{B}$ a Banach space, we let the *strong topology* or *norm topology* be generated by taking as our subbasic open sets those of the form $\{v \in \mathbb{B} : \|v - w\| < \epsilon\}$ for some $\epsilon > 0, w \in \mathbb{B}$. The *weak topology* is generated by subbasic open sets of the form $\{v \in \mathbb{B} : |\phi(v) - a| < \epsilon\}$, for some $\epsilon > 0, a \in \mathbb{K}, \phi \in \mathbb{B}^*$.

These two definitions can in turn be applied to $\mathbb{B}^*$. We have its norm topology and we have its weak topology, obtained by looking at all elements of $\mathbb{B}^{**}$, the dual of $\mathbb{B}^*$. Here however there is a third topology.

We define the *weak* topology* on $\mathbb{B}^*$ by taking as our subbasic open sets all the sets of the form

$$\{\phi \in \mathbb{B}^* : |\phi(v) - a| < \epsilon\}.$$

Definition Thus it is the topology on $\mathbb{B}^*$ generated by the basic open sets of the form

$$\{\varphi \in \mathbb{B}^* : \varphi(z_0) \in V_0, \varphi(z_1) \in V_1, ...\varphi(z_n) \in V_n\}$$

for $z_0, ..., z_n \in \mathbb{B}$, $V_0, V_1, ..., V_n$ open subsets of $\mathbb{K}$. We say that a set is *weak * open*, *weak* closed*, or *weak* compact* when it is respectively open, closed, or compact in this topology.

As a warning on notation, many people use "*weak**" or "weak star" to denote this topology.

Exercise (Alaoglu's theorem) The *unit ball of* $\mathbb{B}^*$ consists of all φ for which

$$|\varphi(z)| \leq 1$$

for every $z \in \mathbb{B}$ with $\|z\| \leq 1$. Show that the unit ball of $\mathbb{B}^*$ equipped with the weak star topology is compact.
(Hint: This space is a closed subset of

$$\prod_{z\in\mathbb{B},\|z\|\leq 1}[-1,1],$$

the collection of all functions from the unit ball of $\mathbb{B}$ to $[-1,1]$, and so we can apply Tychonov's theorem. See V§3 of [3] for more details.)

The critical point of the weak* topology is that it frequently admits metrization.

Theorem 10.1 *Let $\mathbb{B}$ be a separable Banach space. Then $(\mathbb{B}^*)_1$ in the weak* topology admits a complete compatible metric.*

I am not going to enter into the details of this proof, but I will describe the construction. Let $\{v_i : i \in \mathbb{N}\}$ be dense in $\mathbb{B}$ (with respect to the norm topology – that is to say, for each $w \in \mathbb{B}$ and $\epsilon > 0$ there exists i with $\|v_i - w\| < \epsilon$). Then we can take as our metric

$$d(\varphi_1, \varphi_2) = \sum_{i\in\mathbb{N}} \min\{2^{-i}, |\varphi_1(v_i) - \varphi_2(v_i)|\}.$$

Theorem 10.2 *(*The Hahn-Banach theorem*) Let V be a vector space over $\mathbb{K}$ ($\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$). Let*

$$q : V \to \mathbb{R}^{\geq 0}$$

be a function which is sublinear *in the sense that:*

1. *$q(v + w) \leq q(v) + q(w)$ all $v, w \in V$, and*
2. *$q(cv) = cq(v)$ for $v \in \mathbb{B}$, $c \geq 0$.*

Let V_0 be a subspace of V and $\phi_0 : V \to \mathbb{K}$ linear with $|\phi_0(v)| \leq q(v)$ all $v \in V_0$.
Then there exists a linear function $\phi : V \to \mathbb{K}$ with $\phi|_{V_0} = \phi_0$ and $|\phi(v)| \leq q(v)$ all $v \in V$.

See III.6 [3] or 9.4 [7].
A specific case of this theorem is in the situation that $V = \mathbb{B}$ is a Banach space, $q(\cdot) = \|\cdot\|$, and $V_0 = \{cv_0 : c \in \mathbb{K}\}$, some $v_0 \in \mathbb{B}$ with $\|v\| = 1$, and

$$\phi_0 : V_0 \to \mathbb{K},$$

$$cv_0 \mapsto c.$$

Then we can apply 10.2 to ϕ_0 and the sublinear $q(v) = \|v\|$ to obtain $\phi \supset \phi_0$ with ϕ linear, defined on all of $\mathbb{B}$, and with the bound in norm

$$|\phi(v)| \leq \|v\|.$$

This in particular gives that $\mathbb{B}^*$ is rich enough to separate points. For any $v \in \mathbb{B}$,

$$\|v\| = \sup\{|\phi(v)| : \phi \in \mathbb{B}^*, \|\phi\| \leq 1\}.$$

A further consequence of the Hahn-Banach theorem relates to the ability of $\mathbb{B}^*$ to separate a point from a closed subset.

Theorem 10.3 *Let $\mathbb{B}$ be a Banach space and $\mathbb{B}_0$ a closed subspace. Let $v \in \mathbb{B}$ with $v \notin \mathbb{B}_0$. Then there exists $\phi \in \mathbb{B}^*$ with $\phi(v) \neq 0$ but $\phi(w) = 0$ all $w \in \mathbb{B}_0$.*

Proof We begin by defining the function

$$q : u \mapsto d(u, \mathbb{B}_0),$$

where $d(u, \mathbb{B}_0)$ equals the infinum of the set $\{\|u - w\| : w \in \mathbb{B}_0\}$. It is routine to verify that this function is sublinear, and since $0 \in \mathbb{B}_0$ we have $q(u) \leq \|v\|$.

Now for $v \notin \mathbb{B}_0$ let $\phi_0(v) = q(v)$, and extend this to a linear function on $\mathbb{K} \cdot v$ by $\phi_0(cv) = cq(v)$. Then we obtain ϕ by applying the Hahn-Banach theorem to the sublinear function q and the function

$$\phi_0 : \mathbb{K} \cdot v \to \mathbb{K}.$$

□

There is also a geometric version of Hahn-Banach which is proved using a clever choice of the sublinear q. I will state it just for Banach spaces.

Theorem 10.4 *Let $\mathbb{B}$ be Banach space and $C \subset \mathbb{B}$ a closed subset which is* convex *in the sense that for $v, w \in C, \alpha \in [0, 1]$, we have $\alpha v + (1 - \alpha)w \in C$. Let $w \in \mathbb{B} \setminus C$. Then w can be separated from C by an element of $\mathbb{B}^*$ – more precisely, there will be $\phi \in \mathbb{B}^*$, $c \in \mathbb{R}$, $\epsilon > 0$, such that for all $v \in C$*

$$\mathrm{Re}(\phi(w)) < c < c + \epsilon < \mathrm{Re}(\phi(v)).$$

Here Re ξ refers to the real part of $\xi \in \mathbb{C}$. See IV.3[3] for a proof. Part of the significance and usefulness of this consequence of the Hahn-Banach theorem is that a set of the form

$$\{v \in \mathbb{B} : \|v - w_0\| < \epsilon\}$$

will always be convex, and hence the topology on $\mathbb{B}$ has a basis consisting of convex open sets.

11 The Riesz representation theorem

This section will closely follow the treatment in [10].

Definition Let (K, d) be a compact metric space. $C(K, \mathbb{R})$ consists of all continuous functions

$$f : K \to \mathbb{R}.$$

For $f \in C(K, \mathbb{R})$ we let $\|f\|$ be

$$\sup_{x \in K} |f(x)|.$$

Given $r \in \mathbb{R}$ and $f \in C(K, \mathbb{R})$ we define rf by pointwise multiplication,

$$(rf)(x) = r((f(x)),$$

and similarly for $f, g \in C(K, \mathbb{R})$ we define $f + g$ by pointwise addition,

$$(f + g)(x) = f(x) + g(x).$$

There is a whole train of remarks set into motion by these simple definitions.

First of all, any continuous function on a compact metric space is bounded, so we are indeed assured that $\|f\|$ will always be a real number – the sup does not attain $+\infty$. Given the norm we can define $\rho(f, g) = \|f - g\|$. It is not hard to verify this satisfies the triangle inequality, and then that $\rho(\cdot, \cdot)$ defines a metric.

Secondly, given a sequence $(f_n)_n$ in $C(K, \mathbb{R})$ which is Cauchy with respect to ρ we can easily check that $(f_n(x))_n$ is Cauchy for any x and we can define $f : K \to \mathbb{R}$ by simply letting $f(x)$ be the limit of $(f_n(x))_n$. By considering their being Cauchy with respect to ρ we have $(f_n)_n$ converges to f not just pointwise but also *uniformly*:

$$\forall \epsilon > 0 \exists N \forall x \in K \forall n > N (|f(x) - f_n(x)| \leq \epsilon).$$

(Take N to be large enough such $\forall n, m > N \forall x \in K(|f_n(x) - f_m(x)| < \epsilon$.) It is a standard result and a fairly routine calculation that the pointwise limit of a uniformly convergent sequence of uniformly continuous functions is again uniformly continuous, and thus $f \in C(K, \mathbb{R})$.

In conclusion then we have that $C(K, \mathbb{R})$ is a Banach space.

And given a Banach space we can sensibly ask for its dual. The Riesz representation theorem states that the positive elements of $C(K, \mathbb{R})$ with norm 1 may be identified with the collection of probability measures on K.

We will prove this. It is a long proof.

Definition For K a compact metric space, $C(K, \mathbb{R})^*$, the *dual space*, is the collection of functions

$$\Lambda : C(K, \mathbb{R}) \to \mathbb{R}$$

that are continuous and *linear*: $\Lambda(f + g) = \Lambda(f) + \Lambda(g)$, $\Lambda(rf) = r\Lambda(f)$. We then let $\|\Lambda\|$ be

$$\sup_{\{f \in C(K,\mathbb{R}) : \|f\| = 1\}} |\Lambda(f)|.$$

(It is a routine exercise to verify $\|\Lambda\|$ finite from Λ continuous.)

We say that $f \in C(K, \mathbb{R})$ is *positive* if $f(x) \geq 0$ all $x \in K$. We then write $f \leq g$ if $g - f$ is positive. For $r \in \mathbb{R}$ we write $f \leq r$ (respectively $r \leq f$) if $f(x) \leq r$ (respectively $r \leq f(x)$) all $x \in K$. For $r \in \mathbb{R}$ we abuse notation and also use r to indicate the continuous function

$$r : K \to \mathbb{R},$$

$$x \mapsto rx.$$

We say that $\Lambda \in C(K, \mathbb{R})^*$ is *positive* if $\Lambda(f) \geq 0$ whenever f is positive.

Definition Let K be a compact metric space and $f \in C(K, \mathbb{R})$. The *support* of f is the closure of the set of $x \in K$ with $f(x) \neq 0$. For $C \subset K$ closed,

$$C \prec f$$

indicates $0 \leq f \leq 1$ and $f(x) = 1$ for all $x \in C$. For $V \subset K$ open,

$$f \prec V$$

indicates

$$\overline{\{x : f(x) \neq 0\}} \subset V;$$

in other words, V includes the *support* of f.

Lemma 11.1 *If K is a compact metric space and $C \subset V$ with C closed and V open, then there is a $f \in C(K, \mathbb{R})$ with $0 \leq f \leq 1$ and*

$$C \prec f \prec V.$$

Proof Assume $d(C, K \setminus V) = \epsilon > 0$. Let

$$g(x) = \frac{1}{\epsilon} d(x, K \setminus V)$$

and

$$f(x) = \min(g(x), 1).$$

$\square$

In many cases a measure is defined by an outer measure on the open sets. For instance, with Lebesgue measure, we have that the Lebesgue measure $m(A)$ of a set A is equal to the infinum of $m^*(U)$ as U ranges over open sets including A. In this situation there is a neat test for measurability.

Lemma 11.2 *Let X be a Polish space and let $\mathcal{O}(X)$ be the open subsets of X. Let*

$$\rho : \mathcal{O}(X) \to \mathbb{R}^{\geq 0} \cup \{\infty\}$$

be a function. Let

$$\lambda : \mathcal{P}(X) \to \mathbb{R}^{\geq 0} \cup \{\infty\}$$

be the outer measure defined by

$$\lambda(A) = \inf\{\sum_{n \in \mathbb{N}} \rho(O_n) : \{O_n : n \in \mathbb{N}\} \subset \mathcal{O}(X), A \subset \bigcup_{n \in \mathbb{N}} O_n\}.$$

Then $B \subset X$ is λ-measurable if for every open set O we have

$$\lambda(B) \geq \lambda(B \cap O) + \lambda(B^c \cap O).$$

Proof First of all, the implicit statement in the lemma that λ will define an outer measure is justified by 3.6.

The definition of λ-measurable is that for every $A \subset X$ we have

$$\lambda(A) = \lambda(B \cap A) + \lambda(B^c \cap A).$$

It is immediate from the definition of the outer measure that we always have $\lambda(A) \leq \lambda(B \cap A) + \lambda(B^c \cap A)$, so let us instead concentrate on showing $\geq$.

But given any open set O covering A, we have

$$\lambda(O) = \lambda(B \cap O) + \lambda(B^c \cap O) \geq \lambda(B \cap A) + \lambda(B^c \cap A).$$

Letting

$$\lambda(O) \to \lambda(A)$$

we finish. $\square$

Notation For X a topological space, $\mathcal{K}(X)$ denotes the collection of compact subsets of X.

Definition A topological space X is said to be *locally compact* if every point is included in some open set whose closure is compact. For X a locally compact metric space,

$$\rho : \mathcal{K}(X) \to \mathbb{R}^{\geq 0}$$

is a *content* if:

1. $\rho(K_1 \cup K_2) = \rho(K_1) + \rho(K_2)$ for K_1, K_2 disjoint;
2. $\rho(K_1) \leq \rho(K_2)$ whenever $K_1 \subset K_2$.

Theorem 11.3 *Let X be a locally compact Polish space. Let*

$$\rho : \mathcal{K}(X) \to \mathbb{R}^{\geq 0}$$

be a content. Then there is a measure μ on the Borel subsets of X with the properties that:

1. *$\rho(K) \leq \mu(K)$ all $K \in \mathcal{K}(X)$;*
2. *$\mu(U) \leq \rho(K)$ all $K \in \mathcal{K}(X)$ and open $U \subset K$.*

Proof We define an outer measure λ. As a first step to defining λ we give its precursor λ^* on open sets. For $O \subset X$ open we let

$$\lambda^*(O) = \sup\{\rho(K) : K \in \mathcal{K}(X), K \subset O\}.$$

We then obtain an outer measure with

$$\lambda(A) = \inf\{\lambda^*(O) : O \text{ open}, A \subset O\}.$$

It follows from the logical structure of the definitions that λ extends λ^*. It also follows from the logical structure of the definitions that if U is an open subset of a compact set K then

$$\lambda(U) \leq \rho(K) \leq \lambda(K).$$

This will mean that the measure induced by λ, as described at 3.2, will be as required once we show that the Borel sets are all λ-measurable.

Claim: Every element of $\mathcal{K}(X)$ is λ-measurable.
Proof of Claim: Fix K a compact subset of X. Let $O \subset X$ be open . By 11.2 it suffices to show that

$$\lambda(O) \geq \lambda(O \cap K) + \lambda(O \cap K^c).$$

Note first of all, if K_1 is a compact subset of $O \cap K^c$ and K_2 a compact subset of $O \cap K_1^c$ then they are necessarily disjoint, and hence

$$\lambda(O) \geq \rho(K_1 \cup K_2) = \rho(K_1) + \rho(K_2)$$

by the additivity assumption on ρ. Thus for any compact $K_1 \subset O \cap K^c$ we have

$$\lambda(O) \geq \rho(K_1) + \sup\{\rho(K_2) : K_2 \subset O \cap K_1^c, K_2 \in \mathcal{K}(X)\}$$

$$= \rho(K_1) + \lambda(O \cap K_1^c),$$

which in turn at least equals $\rho(K_1) + \lambda(O \cap K)$, since $K_1 \subset K^c$ implies $K_1^c \supset K$. In other words, we have established that for any compact $K_1 \subset O \cap K^c$,

$$\lambda(O) \geq \lambda(O \cap K) + \rho(K_1).$$

Hence,

$$\lambda(O) \geq \lambda(O \cap K) + \sup\{\rho(K_1) : K_1 \subset O \cap K^c, K \in \mathcal{K}(X)\}.$$

But $O \cap K^c$ is open, and thus

$$\lambda(O \cap K^c) = \sup\{\rho(K_1) : K_1 \subset O \cap K^c, K \in \mathcal{K}(X)\},$$

and completes the proof of the claim. (Claim□)

Having established that the compact subsets of X are all λ-measurable, it suffices by 3.2 to see that any σ-algebra containing the compact sets contains the open sets, but this follows from the space being locally compact. □

Lemma 11.4 *If K is a compact metric space and $V_1, ..., V_n$ are open sets with*

$$\bigcup_{i \leq n} V_i \supset C,$$

C closed, then there are continuous functions $h_1, ..., h_n$ with

$$0 \leq h_i \leq 1,$$

$$h_i \prec V_i,$$

at each i and

$$h_1(x) + h_2(x) + ... + h_n(x) = 1$$

at each $x \in C$.

Proof For each $x \in C$ we can find an open neighborhood U_x such that

$$\overline{U_x} \subset V_i$$

for some i. By compactness we may cover C with finitely many of these sets of the form U_x; call them $O_1, O_2, ..., O_\ell$. At $i \leq n$ we let C_i be the union of all $\overline{O_j}$'s with $\overline{O_j} \subset V_i$; this set is a finite union of closed sets and hence closed. Applying 11.1 we may find continuous g_i with $C_i \prec g_i \prec V_i$. The C_i's cover C and hence every $x \in C$ has some i with $g_i(x) = 1$, and so

$$(1 - g_1)(1 - g_2)...(1 - g_n)(x) = 0.$$

Thus if we let

$$h_1 = g_1,$$

$$h_2 = (1 - g_1)g_2,$$

$$h_3 = (1 - g_1)(1 - g_2)g_3,$$

through to

$$h_n = (1 - g_1)(1 - g_2)...(1 - g_{n-1})g_n$$

then at each j

$$h_1 + h_2 + ... + h_j = 1 - (1 - g_1)(1 - g_2)...(1 - g_j)$$

yields

$$h_1 + h_2 + ... + h_j + h_{j+1} = 1 - ((1 - g_1)(1 - g_2)...(1 - g_j) - (1 - g_1)(1 - g_2)...(1 - g_j)g_{j+1})$$

$$= 1 - (1 - g_1)(1 - g_2)...(1 - g_j)(1 - g_{j+1}),$$

until at last

$$h_1 + h_2 + ... + h_n = 1 - (1 - g_1)(1 - g_2)...(1 - g_n),$$

which by the above constantly assumes the value 1 on C. □

Theorem 11.5 *Let K be a compact metric space and suppose $\Lambda \in C(K, \mathbb{R})^*$ is positive with $\|\Lambda\| = 1$. Then there is a Borel probability measure μ on K with*

$$\Lambda(f) = \int f(x) d\mu(x)$$

for any $f \in C(K, \mathbb{R})$.

Proof For $C \subset K$ closed and $(f_n)_{n\in\mathbb{N}}$ a sequence of elements in $C(K, \mathbb{R})$ we write

$$f_n \to C$$

if:

1. each $f_n|_C \equiv 1$;
2. each f_n has its range included in $[0, 1]$;
3. for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n \geq N$ the support of f_n is included in the set $\{x \in K : d(x, C) < \epsilon\}$.

Claim 1: If $f_n \to C$ then $(\Lambda(f_n))_{n\in\mathbb{N}}$ converges.

Proof of claim: Suppose the $(\Lambda(f_n))_{n\in\mathbb{N}}$ sequence is not Cauchy. In particular fix some $\delta > 0$ such that for all N there exists $n, m \geq N$ with

$$|\Lambda(f_n) - \Lambda(f_m)| > \delta.$$

Then we may find a subsequence $(f_{n(k)})_{k\in\mathbb{N}}$ such that each

$$\Lambda(f_{n(k)}) - \Lambda(f_{n(k+1)}) > \delta/2$$

and the support of $f_{n(k+1)}$ is included in the set

$$\{x \in K : f_{n(k)} \geq 1 - \frac{\delta}{2}\}.$$

(Note: If we ensure that $f_{n(k+1)} < \delta/2 + f_{n(k)}$ then the positivity of Λ entails that the only way we can have $|\Lambda(f_{n(k)}) - \Lambda(f_{n(k+1)})| > \delta/2$ is to actually have $\Lambda(f_{n(k)}) - \Lambda(f_{n(k+1)}) > \delta/2$.) For $k < \ell$ we then have

$$\|f_{n(k)} - f_{n(\ell)}\| < 1 + \frac{\delta}{2},$$

whilst the additivity of Λ yields

$$\Lambda(f_{n(k)}) - \Lambda(f_{n(\ell)}) = \sum_{i<k-\ell} \Lambda(f_{n(k+i)}) - \Lambda(f_{n(k+i+1)}) \geq \frac{\ell\delta}{2},$$

which by letting $\ell \to \infty$ contradicts the boundedness of the operator Λ. (Claim$\square$)

Note then from this claim that given a fix closed $C \subset K$ and two sequences

$$f_n \to C,$$

$$g_n \to C$$

we must have

$$\lim_{n\to\infty} \Lambda(f_n) = \lim_{n\to\infty} \Lambda(g_n).$$

This justifies a definition.

Definition For $C \subset K$ closed let $\rho(C)$ be equal to

$$\lim_{n\to\infty} \Lambda(f_n)$$

for any $f_n \to C$.

Claim 2: If C_1, C_2 are disjoint closed subsets of K, then

$$\rho(C_1 \cup C_2) = \rho(C_1) + \rho(C_2).$$

Proof of Claim: Choose U_1, U_2 disjoint supersets of C_1, C_2 respectively. Choose

$$f_n \to C_1,$$

$$g_n \to C_2,$$

with the support of each f_n included in U_1 and of each g_n included in U_2. It then follows that

$$f_n + g_n \to C_1 \cup C_2$$

and the additivity of Λ gives each $\Lambda(f_n + g_n) = \Lambda(f_n) + \Lambda(g_n)$. (Claim$\square$)

It is easily seen that $C_1 \subset C_2$ yields $\rho(C_1) \leq \rho(C_2)$. Thus we obtain that ρ is a content.

Claim 3: Let C be a closed subset of K. At each n let

$$C_n = \{x \in K : d(x, C) \leq \frac{1}{n}\}.$$

Then $\rho(C_n) \to C$.

Proof of Claim: We can choose a sequence of functions

$$f_n \to C$$

where

$$|\Lambda(f_n) - \rho(C_n)| < \frac{1}{n}.$$

(Claim□)

Thus if we use 11.3 to induce a Borel probability measure μ from ρ, then we will have

$$\mu(C) = \rho(C)$$

any closed $C \subset K$. There is one last garrison of resistance: We still need to show the equality of integration by μ and application of Λ.

Claim 4: If C is closed and f positive with $f(x) \geq r$ on C, then

$$\Lambda(f) \geq r\mu(C).$$

Proof of Claim: Rescaling by and applying linearity, we may assume $r = 1$. Then this follows from the definition of ρ and the fact that μ in this case must extend ρ. (Claim□)

Claim 5: For $f \in C(K, \mathbb{R})$ with $f \geq 0$,

$$\int f d\mu \leq \Lambda(f).$$

Proof of Claim: Choose $\epsilon > 0$ and $y_0 < y_1 < ...y_n$ with each $y_{i+1} - y_i < \epsilon/2$ and the range of f included in the open interval (y_0, y_n). Let E_i be the set on which $f(x)$ lies in the semi open interval $(y_{i-1}, y_i]$. Let $V_i \supset E_i$ be open and $C_i \subset E_i$ closed with

$$\mu(V_i \setminus C_i) < \frac{\epsilon}{2ny_n},$$

and hence

$$\sum \mu(V_i \setminus C_i) y_n < \frac{\epsilon}{2}.$$

Let

$$U_i = V_i \setminus \bigcup_{j \neq i} C_j.$$

The U_i's cover K so we can apply 11.4 to obtain $h_1, h_2, ..., h_n$ with

$$h_1 + h_2 + ... + h_n = 1,$$

each h_j having support in U_j. Since $U_i \cap C_j = 0$ for $i \neq j$, each h_j assumes the constant value 1 on the corresponding C_j. Letting $f_i = h_i f$ we have

$$f = \sum_{i \leq n} f_i$$

$$\therefore \sum_{i<n} y_{i+1}\mu(V_{i+1}) \geq \int f d\mu.$$

Applying this and claim 3 we obtain

$$\Lambda(f) = \sum_{i<n} \Lambda(f_{i+1}) \geq \sum_{i<n} y_i \mu(C_{i+1}).$$

Thus

$$\int f d\mu - \Lambda(f) \leq (\sum_{i<n} \mu(V_{i+1} \setminus C_{i+1}) y_{i+1}) + (\sum_{i<n} \mu(C_{i+1})(y_{i+1} - y_i)$$

$$< y_n \sum_{i<n} \mu(V_{i+1} \setminus C_{i+1}) + (\sum_{i<n} \mu(C_{i+1}))\frac{\epsilon}{2} < \epsilon.$$

(Claim□)

Claim 6: For any $f \in C(K, \mathbb{R})$

$$\int f d\mu \leq \Lambda(f).$$

Proof of Claim:
For r a constant

$$\Lambda(r) = \int r d\mu = r.$$

Thus if we choose r a sufficiently large positive real to ensure $f + r \geq 0$ we can apply the last claim to obtain

$$\Lambda(f + r) \geq \int (f + r) d\mu.$$

Since $\Lambda(f + r) = \Lambda(f) + \Lambda(r) = \Lambda(f) + r$ and since $\int (f + r) d\mu = \int f d\mu + \int r d\mu = \int f d\mu + r$ we are done. (Claim□)

Our victory is almost complete. Replacing f by $-f$ we can as well obtain

$$\int f d\mu \geq \Lambda(f).$$

And it is done and done. □

Corollary 11.6 *If $\Lambda \in C(K, \mathbb{R})^*$ is positive, then there is a finite Borel measure μ with*

$$\Lambda(f) = \int f d\mu.$$

Proof Apply the last theorem to

$$f \mapsto \frac{\Lambda(f)}{\Lambda(1)}.$$

□

Notation For K a compact metric space, let $P(K)$ be the probability measures on K equipped with the topology generated by the basic open sets

$$\{\mu : s_1 < \mu(f_1) < r_1, s_2 < \mu(f_2) < r_2, ..., s_n < \mu(f_n) < r_n\}$$

for $f_1, f_2, ..., f_n \in C(K, \mathbb{R})$.

Exercise (i) Let K be a compact metric space. Let $C(K, [-1, 1])$ be the subspace of $C(K, \mathbb{R})$ consisting of continuous functions with norm at most one – that is to say, the range included in $[1, -1]$ Show that if $\{f_i : i \in \mathbb{N}\}$ is a countable dense subset of $C(K, [-1, 1])$ then the function

$$\pi : P(K) \to \prod_{i \in \mathbb{N}} [0, 1]$$

given by

$$(\pi(\mu))(n) = \mu(f_n)$$

is continuous and open onto its image (i.e. π effects a homeomorphism between $P(K)$ and $\pi[P(K)]$).

(i) Show that $P(K)$ is a compact metrizable space.

The result also extends to signed measures, and gives a complete description of the dual. To head off any confusion between the real valued dual and the complex dual, keep in mind that for us the dual $C(K, \mathbb{R})$ is the collection of linear, bounded, *real valued* functions.

Lemma 11.7 *Let $\Lambda \in C(K, \mathbb{R})^*$. Then there is a positive $\Phi \in C(K, \mathbb{R})^*$ with*

$$\Phi(f) \geq |\Lambda(f)|$$

for all $f \geq 0$.

Proof For $f \geq 0$ set $\Phi^+(f)$ to be

$$\sup\{|\Lambda(g)| : |g| \leq f\}.$$

Claim: Φ^+ is linear on its domain, in the sense that for $f_1, f_2, r \geq 0$

$$\Phi^+(f_1 + f_2) = \Phi^+(f_1) + \Phi^+(f_2)$$

and

$$\Phi^+(rf_1) = r\Phi^+(f_1).$$

Proof of Claim: The preservation with respect to multiplication by positive scalars is pretty immediate. The main issue is verifying the sups involved in finite additivity.

To see $\Phi^+(f_1 + f_2) \geq \Phi^+(f_1) + \Phi^+(f_2)$, consider some g_1, g_2 with $|g_i| \leq f_i$. After possibly replacing g_i by $-g_i$ we can assume $\Lambda(g_i) \geq 0$, when $|\Lambda(g_1 + g_2)| = |\Lambda(g_1)| + |\Lambda(g_2)|$. Conversely if $|g| \leq f_1 + f_2$, we can again assume without loss of generality that $\Lambda(g) \geq 0$ and write

$$g = g^+ - g^-,$$

where $g^+, g^- \geq 0$, $|g| = |g^+| + |g^-|$. Let $g_1^+ = \min(g^+, f_1)$, $g_1^- = \min(g^-, f_1)$, and then $g_2^+ = g^+ - g_1^+$, $g_2^- = g^- - g_1^-$. So we have found g_1, g_2 with $|g_i| \leq f_i$, $g_1 + g_2 = g$. (Claim$\square$)

Then given any $f \in C(K, \mathbb{R})$ we can let f^+ be defined by $f^+(x) = f(x)$ if $f(x) \geq 0$ and $f^+(x) = 0$ if $f(x) < 0$. Similarly we can let f^- be defined by $f^-(x) = -f(x)$ if $f(x) \leq 0$ and $f^+(x) = 0$ if $f(x) > 0$. Therefore we have represented f as the difference of two continuous functions:

$$f(x) = f^+(x) - f^-(x).$$

We let $\Phi(f) = \Phi^+(f^+) - \Phi^+(f^-)$.

Claim: If $f = g_1 - g_2$ with $g_1, g_2 \geq 0$, then $\Phi(f) = \Phi^+(g_1) - \Phi^+(g_2)$.

Proof of Claim: Let $g(x) = \min(g_1(x), g_2(x))$. Then $f^+(x) = g_1(x) - g(x)$ and $f^-(x) = g_2(x) - g(x)$. Thus by linearity of Φ^+ we have

$$\Phi(f) = \Phi^+(f^+) - \Phi^+(f^-)$$
$$= \Phi^+(g_1) - \Phi^+(g) - (\Phi^+(g_2) - \Phi^+(g)) = \Phi^+(g_1) - \Phi^+(g_2).$$

(Claim□)

From this we have we can routinely verify linearity: For instance,

$$\Phi(f_1 + f_2) = \Phi(f_1^+ + f_2^+ - f_1^- - f_2^-) =$$
$$\Phi^+(f_1^+ + f_2^+) - \Phi^+(f_1^- + f_2^-) = \Phi^+(f_1^+) + \Phi^+(f_2^+) - \Phi^+(f_1^-) - \Phi^+(f_2^-)$$
$$= \Phi(f_1) + \Phi(f_2).$$

Φ provides a positive element of the dual, and the very way in which it has been defined gives $\Phi(f) \geq |\Lambda(f)|$ for all $f \geq 0$. □

Theorem 11.8 *If $\Lambda \in C(K, \mathbb{R})^*$ then there is a signed measure ν with*

$$\Lambda(f) = \int f d\nu.$$

Proof Obtain Φ as in 11.7 and then let μ be as in 11.6 with

$$\Phi(f) = \int f d\mu.$$

Note then that $\int |f| d\mu < r$ entails $\Lambda(f) < r$ and so Λ defines a bounded linear function on the continuous functions in $L^1(\mu)$. The continuous functions are dense in $L^1(\mu)$ and so Λ extends uniquely to the dual of $L^1(\mu)$.

The dual of L^1 is L^∞ (see for instance §B [3] or 17.4.4 [7]) and so we can find $\phi \in L^\infty(\mu)$ with

$$\Lambda(f) = \int f \cdot \phi d\mu.$$

ν defined by

$$\nu(B) = \int_B \phi d\mu$$

is as required. □

There is also a version of this result for the *complex* dual of the continuous functions from K to $\mathbb{C}$, $C(K, \mathbb{C})$: For every $\Lambda \in C(K, \mathbb{C})^*$ there is a finite complex valued measure μ with $\Lambda(f) = \int f d\mu$. This can be easily derived from the previous results.

Give $\Lambda \in C(K, \mathbb{C})^*$ and $f : K \to \mathbb{R}$ continuous, we can write

$$\Lambda(f) = \Lambda_0(f) + i\Lambda_1(f),$$

where $\Lambda_0(f), \Lambda_1(f) \in \mathbb{R}$. It is easily seen that Λ_0, Λ_1 are linear and so we obtained signed, real valued measures μ_0, μ_1 with

$$\Lambda(f) = \int f d\mu_0 + i \int f \mu_1$$

all real valued f. By linearity of integration and Λ, the same formula holds for all complex valued continuous functions. Thus:

Theorem 11.9 *Let K be a compact metric space. Let Λ be a continuous linear function from $C(K, \mathbb{C})$ (the continuous complex valued functions on K) to $\mathbb{C}$.*

Then there is a complex valued Borel measure μ with

$$\Lambda(f) = \int f d\mu$$

for all continuous $f : K \to \mathbb{C}$.

12 Stone-Weierstrass

As in the theorems from the last section, I will be taking the Banach spaces and vector spaces over $\mathbb{R}$. There is a parallel sequence of results for complex functions, but I do not want to clutter the path with diversions.

Definition A subset A of a vector V is said to be *convex* if for all $a, b \in A$ and $\alpha \in [0,1]$ we have

$$\alpha a + (1-\alpha)b \in A.$$

$c \in A$ is then said to be an *extreme point of* A if whenever $\alpha \in (0,1)$, $a, b \in A$ with

$$\alpha a + (1-\alpha)b = c$$

we have $a = b = c$.

Exercise (i) Let $\mathbb{B}$ be a Banach space. Show that the basic open sets in the weak* topology on $\mathbb{B}^*$ are all convex.

(ii) Show that if $V \subset \mathbb{B}^*$ is convex and weak* open, then for any $\phi \in \mathbb{B}^*$ the set

$$V + \phi = \{\psi + \phi : \psi \in V\}$$

is also convex and weak* open as is

$$-V = \{-\psi : \psi \in V\}.$$

Lemma 12.1 *Let $\mathbb{B}$ be a Banach space and $A \subset \mathbb{B}^*$ convex and and weak* open. Let ϕ be in the weak star closure of A. Then for any $\psi \in A$ and $t \in (0,1)$ we have*

$$t\phi + (1-t)\psi \in A.$$

Proof Let V be a convex, weak * open neighborhood of the identity with $\psi + V \subset A$. Note that

$$t^{-1}(1-t)V =_{\text{df}} \{t^{-1}(1-t)\varphi : \varphi \in V\}$$

is again an open neighborhood of 0, and so

$$\phi - t^{-1}(1-t)V$$

is an open neighborhood of ϕ, and thus we can find some $\theta \in A$ lying in this set, which amounts to

$$\phi - \theta \in t^{-1}(1-t)V.$$

This in turn yields

$$t\phi \in t\theta + (1-t)V$$

$$\therefore t\phi + (1-t)\psi \in t\theta + (1-t)(\psi + V).$$

Choosing some $\zeta \in V$ with

$$t\phi + (1-t)\psi = t\theta + (1-t)(\psi + \zeta),$$

we have $\psi + \zeta \in A$ since $\psi + V \subset A$, and then $t\phi + (1-t)\psi$ is a convex linear combination of elements in A, as required. □

Theorem 12.2 *(Krein-Milman) Let $\mathbb{B}$ be a Banach space and let $A \subset \mathbb{B}^*$ satisfy:*

(i) A non-empty;

(ii) A convex;

(iii) A compact in the weak topology.*

Then A contains an extreme point.

Proof Let $\mathcal{U}$ be the collection of all convex proper subsets of A which are relatively open in the weak star topology. The compactness of A ensures that $\mathcal{U}$ is closed under directed unions, and so we can apply Zorn's lemma to get a maximal element, V. It suffices to show $A \setminus V$ consists of a singleton.

Let $\overline{V}$ be the weak star closure of V.

Claim: $V \neq \overline{V}$.

Proof of Claim: Choose any $\psi \in V, \phi \notin V$. Let $S = \{s \in [0,1] : s\psi + (1-s)\phi \in V\}$. This set is open in $[0,1]$ since $s \mapsto s\psi + (1-s)\phi \in V$ is weak star continuous. Let s_0 be the infinum of $[0,1] \setminus S$. Then $s_0\psi + (1-s_0)\phi$ is in $\overline{V} \setminus V$. (Claim$\Box$)

Claim: If W is a convex subset of A, then $V \cup W$ is a convex subset of A.

Proof of Claim: It suffices to show that if $\psi \in V$, $\phi \in W$, and $\alpha \in (0,1)$, then $\alpha\psi + (1-\alpha)\phi \in V$.

Define the function

$$T : A \to A$$

$$\varphi \mapsto \alpha\psi + (1-\alpha)\varphi.$$

T is continuous, affine, injective, and has $T[\overline{V}] \subset V$ by 12.1. Hence $T^{-1}[V]$ is a convex open subset of A including $\overline{V}$, and thus by maximality of V and the last claim it must equal A. In particular $T(\phi) \in V$. (Claim$\Box$)

Now for a contradiction, assume ϕ_0, ϕ_1 are distinct points in $A \setminus V$. Let W_0 be a convex open set containing the first point but not the second. Then $V \cup W_0$ is a convex set, weak star open set providing a counterexample to the maximality of V. $\Box$

The version of Krein-Milman presented above is rather weak. First of all, the proper conclusion of the result is not just that A contains an extreme point, but moreover the convex closure of the extreme points is dense in A. Secondly, the result holds in greater generality: We only really used that $\mathbb{B}^*$ is a topological vector space with a basis consisting of convex sets.

Definition For K a compact metric space and μ a signed measure, we appeal to the Jordan decomposition of §3 above to find orthogonal, finite (positive) measures μ^+, μ^- with

$$\mu = \mu^+ - \mu^-.$$

We then let $|\mu| = \mu^+ + \mu^-$.

Theorem 12.3 $\|\mu\| = \||\mu|\| = |\mu|(K)$.

Aside:Here $\|\mu\|$ refers to its norm viewed as an element of $C(K,\mathbb{R})^*$, the dual space to $C(K,\mathbb{R})$. That is to say, if we define

$$\Lambda_\mu : C(K,\mathbb{R}) \to \mathbb{R}$$

$$f \mapsto \int f d\mu,$$

then $\|\mu\|$ equals the norm of Λ_μ as an element of $C(K,\mathbb{R})^*$.

Proof Let A be as in the Jordan decomposition, with

$$\mu(A \cap B) \geq 0,$$

$$\mu(A^c \cap B) \leq 0$$

all Borel B and $\mu^+(B) = \mu(B \cap A)$, $\mu^-(B) = \mu(B \cap A^c)$.

$\|\Lambda_\mu\| \geq |\mu|(K)$:-

Fix $\epsilon > 0$. Fixed closed sets C_1, C_2 with

$$C_1 \subset A,$$

$$C_2 \subset A^c,$$

$$|\mu|(K \setminus (C_1 \cup C_2)) < \epsilon.$$

Fix continuous

$$f_1, f_2 : K \to [0, 1]$$

with f_1 constantly $= 1$ on C_1, constantly $= 0$ on C_2, f_2 constantly $= 1$ on C_2, constantly $= 0$ on C_1. Then let $f = f_1 - f_2$.

$$|\Lambda_\mu(f) - \int_{C_1 \cup C_2} f d\mu| = |\int f d\mu - \int_{C_1 \cup C_2} f d\mu| < 2\epsilon,$$

by the assumption on the measure of $C_1 \cup C_2$ and using $|f| \leq 1$. However since $f|_{C_1} \equiv 1$, and $f|_{C_2} \equiv -1$ we have

$$\int_{C_1 \cup C_2} f d\mu = \mu(C_1) - \mu(C_2) = |\mu|(C_1 \cup C_2).$$

Since $|\Lambda_\mu(f) - \int_{C_1 \cup C_2} f d\mu| < 2\epsilon$ and $|\mu(K) - \mu(C_1 \cup C_2) < \epsilon$ we get

$$|\Lambda_\mu(f)| \geq |\mu|(K) - 3\epsilon.$$

$\|\Lambda_\mu\| \leq |\mu|(K)$:-

For the reverse inequality, we have at any f with $\|f\| \leq 1$ that

$$|\Lambda_\mu(f)| = |\int f d\mu| \leq |\int_A f d\mu| + |\int_{A^c} f d\mu|,$$

which is in turn bounded by $\mu^+(A) + \mu^-(A^c) = |\mu|(K)$. □

Definition $\mathcal{A} \subset C(K, \mathbb{R})$ is an *algebra* if it is closed under addition ($f, g \in \mathcal{A} \Rightarrow (x \mapsto f(x) + g(x)) \in \mathcal{A}$), multiplication ($f, g \in \mathcal{A} \Rightarrow (x \mapsto f(x)g(x)) \in \mathcal{A}$) , and scalar multiplication ($f \in \mathcal{A}, r \in \mathbb{R} \Rightarrow (x \mapsto rf(x)) \in \mathcal{A}$).

An algebra $\mathcal{A}$ is said to *separate points* if for all $x_0 \neq x_1$ in K there is some $f \in \mathcal{A}$ with

$$f(x_0) \neq f(x_1).$$

Theorem 12.4 *(Stone-Weierstrass) Let $\mathcal{A} \subset C(K, \mathbb{R})$ be an algebra which is closed in the sup norm, separates points, and contains the constant functions. Then $\mathcal{A} = C(K, \mathbb{R})$.*

Proof Consider the subset A of $C(K, \mathbb{R})^*$ consisting of Λ with norm less than one having

$$\Lambda(f) = 0$$

all $f \in \mathcal{A}$. By Alaoglu's theorem, this is a compact set in the weak topology. It is clearly convex. In light of the Hahn-Banach theorem (as found presented at 9.4.4. [7]) we will be done if we show that A only consists of 0.

Note that if $A \neq \{0\}$ then all of its extreme points would have to have norm one. So, instead, for a contradiction, apply 12.2 to obtain some extreme $\Lambda \in A$, $\|\Lambda\| = 1$. Apply 11.5 to get a signed measure μ with

$$\Lambda(f) = \int f d\mu$$

all $f \in C(K, \mathbb{R})$. By the exercise above we have $|\mu|(K) = 1$. Let C be the *support* of $|\mu|$:

$$C =_{\mathrm{df}} K \setminus \bigcup \{U : U \text{ open }, |\mu|(U) = 0\}.$$

Thus C is the minimal closed set with $\mu(E) = 0$ all Borel $E \subset K \setminus C$.

Claim: C does not consist of a single point.

Proof of Claim: Otherwise suppose $C = \{x_0\}$. Then for any $g \in C(K, \mathbb{R})$

$$\Lambda(g) = \Lambda(c_0),$$

where $c_0 = g(x_0)$. But c_0 is in $\mathcal{A}$ since it is a constant function, and after all we have $\Lambda(g) = 0$ all $g \in C(K, \mathbb{R})$. (Claim$\square$)

Let $x_0 \neq x_1$ be in C. Since $\mathcal{A}$ separates points we obtain some $f_0 \in \mathcal{A}$ with

$$f_0(x_0) \neq f_0(x_1).$$

Using the constant functions in $\mathcal{A}$ along with closure under addition we can get $f_1 \in \mathcal{A}$ with

$$f_1(x_0) = 0,$$

$$f_1(x_1) \neq 0.$$

Letting $f_2 = (f_1)^2$ we obtain a positive element of $\mathcal{A}$ with the same property. Letting

$$f_3 = \frac{f_2}{\|f_2\|}$$

we obtain $f_3 \in \mathcal{A}$ with

$$f_3 : K \to [0, 1],$$

$$f_3(x_0) = 0,$$

$$f_3(x_1) \neq 0.$$

We now define a new measure, $f_3\mu$, with

$$\int f d(f_3\mu) = \int (f \cdot f_3)\mu.$$

$|f_3\mu| = f_3|\mu|$ since $f_3 \geq 0$. Let

$$\alpha = \|f_3\mu\|.$$

Claim: $0 < \alpha < 1$.

Proof of Claim: $0 < \alpha$ since $f_3(x_1) > 0$. Conversely, $f_3(x_0) = 0$ so we can find $f \in C(K, \mathbb{R})$ with

$$f(x_0) \neq 0,$$

$$0 \leq f \leq 1,$$

$$0 \leq f + f_3 \leq 1.$$

Then

$$\int f + f_3|\mu| > \int f_3|\mu|$$

since f takes positive values around x_0 which is in the support of μ. Since $\int f + f_3|\mu| \leq 1$, we are done. (Claim$\Box$)

We have constrained f_3 so it only takes values between 0 and 1, and hence $1 - f_3 \geq 0$, and so

$$\|(1 - f_3)\mu\| = |(1 - f_3)\mu|(K) = \int (1 - f_3)|\mu| = 1 - \alpha.$$

Then

$$\mu = \alpha \frac{f_3\mu}{\|f_3\mu\|} + (1 - \alpha)\frac{(1 - f_3)\mu}{\|1 - f_3\mu\|}.$$

From Λ being an extreme point we have $\mu = f_3\mu/\|f_3\mu\|$, but our assumptions on x_0, x_1 give us open neighborhoods U_0, U_1 of x_0, x_1 with

$$|\mu|(U_0), |\mu|(U_1) \neq 0$$

and some constant c such that for all $z \in U_0, z' \in U_1$

$$f_3(z) < c < f_3(z'),$$

with a contradiction. $\Box$

A similar argument holds for algebras of *complex* valued functions with the sup norm

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

given $f, g : K \to \mathbb{C}$. There are however two key differences.

The first of these differences is minor. In the proof above we needed to use the Riesz representation theorem to summon in to being the indicated measure. In the complex valued case we need the refinement at 11.9.

The second difference is more telling. Trailing through the proof above we reached the function f_1 which separated the points x_0, x_1 as indicated, and from there we passed to $f_2 \geq 0$ which performed a similar task. In the complex value case we cannot simply take $f_2 = (f_1)^2$ since this will not necessarily be real valued.

Instead we must make an additional assumption on the algebra. We need to assume it is closed under *complex conjugation* – which is to say whenever $f \in \mathcal{A}$ we also have in $\mathcal{A}$ the function

$$\bar{f} : x \to \overline{f(x)},$$

mapping x to the complex conjugate of $f(x)$. With this key adjustment we can let

$$f_2 = f_1 \cdot \overline{f_1}$$

and continue the proof as above.

Theorem 12.5 *Let K be a compact metric space and let $\mathbb{C}(K, \mathbb{C})$ be the complex valued continuous functions on K. Let $\mathcal{A} \subset C(K, \mathbb{C})$ be such that:*

(i) $\mathcal{A}$ is an algebra, in the set that $f_1, f_2 \in \mathcal{A}$, $\alpha \in \mathbb{C}$ yield

$$f_1 + f_2 \in \mathcal{A},$$

$$\alpha f_1 \in \mathcal{A};$$

(ii) $\mathcal{A}$ contains the constant complex valued functions;
(iii) $\mathcal{A}$ separates points;
(iv) $\mathcal{A}$ is closed under complex conjugation.

Then $\mathcal{A}$ is dense in $C(K, \mathbb{C})$.

There is an interesting corollary to Stone-Weiestrass simply from the point of view of measurable functions. If X is a metric space and μ is a Borel probability measure on X, then the continuous functions are dense in the measure theoretic Banach spaces $L^p(X, \mu)$, $1 \leq p < \infty$. (This hopefully is familiar to you from 3rd year analysis.) Thus if $\mathcal{A} \subset C(K, \mathbb{R})$ is an algebra of functions which separates points, then $\mathcal{A}$ will be dense viewed as a subset of the Banach space $L^p(X, \mu)$.

13 Product measures

Given two measures μ and ν on X and Y, we can form a new measure $\mu \times \nu$ on the product space. There are two main things we want to establish: Firstly, that this measure makes sense – it can be defined and there is truly only one choice for how we would to do this; secondly, Fubini's theorem: integrating over $\mu \times \nu$ is the same as integrating by μ and then ν or by ν and then μ. It will be helpful in the course of the proofs to use a couple of the basic convergence theorems. You would probably have seen them before, at least in some form, but there is no harm in quickly recalling the proofs.

All the spaces in this section will be standard Borel. All the measures will be non-negative Borel probability measures. The results hold in the more general context of σ-finite Borel measures, but it is usually a routine exercise to obtain the general result from the specific ones we give below.

Notation If $A \subset X \times Y$ and $x \in X$, then

$$A_x = \{y : (x, y) \in A\}.$$

If $y \in Y$ then

$$A^y = \{x : (x, y) \in A\}.$$

Definition If X and Y are standard Borel spaces, then we equip $X \times Y$ with the σ-algebra generated by the rectangles of the form

$$A \times B \subset X \times Y$$

for A Borel in X and B Borel in Y.

Lemma 13.1 *$X \times Y$ in this σ-algebra is a standard Borel space.*

Proof Fix compatible Polish topologies for X and Y. Recall from an earlier exercise that $X \times Y$ is then Polish in the product topology. We finish by observing that every rectangle of the form $A \times B$ as above appears in the σ-algebra generated by the open sets in $X \times Y$. □

Lemma 13.2 *For X, Y as above, and $B \subset X \times Y$ Borel,*

$$B_x$$

and

$$B^y$$

are Borel for all $x \in X$, $y \in Y$.

Proof Observe that the set of B's with these slices Borel forms a σ-algebra containing the open sets. □

Lemma 13.3 *Let (X, μ) and (Y, ν) be standard Borel probability spaces. If $C \subset X \times Y$ is Borel then*

$$x \mapsto \nu(C_x)$$

is measurable.

Proof Actually this takes some work to prove, and I am going to skip through some of the details. (In fact, an even stronger result is true, namely that $x \mapsto \nu(C_x)$ is Borel, but that takes more trouble and would be beyond what we need.)

This is clear for rectangles of the form $A \times B$, A, B Borel subsets of X, Y. We obtain the result for finite unions of rectangles by using that the finite sums of measurable functions are measurable and since every such union can be represented as a finite union of *disjoint* rectangles. From there we can obtain the result for C a countably infinite unions of rectangles by appealing to the monotone convergence theorem. Note that the intersections of rectangles are rectangles, and hence the intersection of two sets formed as the countable unions of rectangles is again a countable union of rectangles.

Claim: If $C \subset X \times Y$ is Borel, then for any $\epsilon > 0$ we can find unions $D^0 = \bigcup_{i \in \mathbb{N}} A_i^0 \times B_i^0$ and $D^1 = \bigcup_{i \in \mathbb{N}} A_i^1 \times B_i^1$ with

$$C \subset D^0,$$
$$X \setminus C \subset D^1,$$

and $\int (x \mapsto \nu(D^0 \cap D^1)_x) d\mu < \epsilon$.

Proof of Claim: Clearly the claim is true for C itself a rectangle of the form $A \times B$, since we can then take $D^0 = A \times B$, $D^1 = (X \setminus A) \times Y \cup X \times (Y \setminus B)$. Clearly if the claim is true for C then it is true for $X \setminus C$. The main battle is to show closure under countable intersections.

For this purpose, suppose the claim is true for $C_1, C_2, C_3, ...$ At each n choose D_n^0, D_n^1 as in claim, with $C_n \subset D_n^0, X \setminus C_n \subset D_n^1$,

$$\int (x \mapsto \nu(D_n^0 \cap D_n^1)_x) d\mu < 2^{-n}\epsilon.$$

Then

$$\int (x \mapsto \nu(\bigcup_{n \in \mathbb{N}} D_n^0 \cap (\bigcap_{n \in \mathbb{N}} D_n^1)_x)) d\mu < \epsilon.$$

By the monotone convergence theorem we eventually get large N with

$$\int (x \mapsto \nu(\bigcup_{n \in \mathbb{N}} D_n^0 \cap (\bigcap_{n \leq N} D_n^1)_x)) d\mu < \epsilon.$$

Since $\bigcap_{n \leq N} D_n^1$ can be expressed as a countable union of rectangles, we have established the claim for $\bigcup_{n \in \mathbb{N}} C_n$.

Since the Borel sets in $X \times Y$ are the smallest σ-algebra containing the rectangles of the form $A \times B$ for $A \subset X$, $B \subset Y$ both Borel, we are done. (Claim$\square$)

We can then obtain for any Borel C two sequence of decreasing sets as above, $(D_n^0)_n, (D_n^1)_n$, with $C \subset D_n^0$, $X \setminus C \subset D_n^1$,

$$\int (x \mapsto \nu((D_n^0)_x \cap (D_n^1)_x)) d\mu \to 0$$

and hence for μ-almost every x

$$\nu((D_n^0)_x \cap (D_n^1)_x) \to 0.$$

Since

$$1 = \nu((D_n^0)_x \cup (D_n^1)_x) = \nu((D_n^0)_x) - \nu((D_n^0)_x \cap (D_n^1)_x) + \nu((D_n^1)_x)$$

we obtain for a.e. x

$$\nu((D_n^0)_x) - (1 - \nu((D_n^1)_x)) \to 0.$$

Since $x \mapsto \nu(C_x)$ is trapped between the measurable functions $x \mapsto (D_n^0)_x$ and $x \mapsto 1 - \nu((D_n^1)_x)$ we obtain that it as well will be measurable. $\square$

Lemma 13.4 *Let (X, μ) and (Y, ν) be standard Borel probability spaces. If we define m on $X \times Y$ by*

$$m(B) = \int_X (x \mapsto \nu(B_x)) d\mu$$

then m is a Borel probability measure on X.

Proof The main issue is to show σ-additivity. So suppose (B_i) is a sequence of disjoint Borel subsets of $X \times Y$. But then at each finite N we have

$$m(\bigcup_{i<N} B_i) = \int_X (x \mapsto \nu((\bigcup_{i<N} B_i)_x)\mu$$

$$= \sum_{i<N} \int (x \mapsto \nu((B_i)_x)\mu = \sum_{i<N} m(B_i).$$

Then the case for $\bigcup_{I\in\mathbb{N}} B_i$ follows by 5.4 or 5.3 and the observation that

$$\nu((\bigcup_{i\in\mathbb{N}} B_i)_x) = lim_{n\to\infty}\nu((\bigcup_{i<n} B_i)_x).$$

□

Lemma 13.5 *Let (X, μ) and (Y, ν) be standard Borel probability spaces. Then for any Borel $B \subset X \times Y$*

$$\int_X (x \mapsto \nu(B_x))d\mu = \int_Y (y \mapsto \mu(B^y))d\nu.$$

Proof Let us first define m as in 13.4 and then likewise m^* by

$$m^*(B) = \int_Y (y \mapsto \mu(B_y))d\nu.$$

This likewise will give us a measure, and we easily check that m and m^* agree on the measurable rectangles. Since every set arising in the algebra generated by the rectangles can be written as a finite union of *disjoint* rectangles, we quickly obtain that these two measures agree on this algebra.

From this it follows (e.g. 15.3.11 of [7], though it is not hard to see directly) that $m = m^*$. □

Definition We let $\mu \times \nu$ be the measure m obtained by either

$$m(B) = \int_X (x \mapsto \nu(B_x))d\mu$$

or

$$m(B) = \int_Y (y \mapsto \mu(B^y))d\nu.$$

Summarising the above:

Proposition 13.6 *Let (X, μ) and (Y, ν) be standard Borel probability spaces. Then the probability measure $\mu \times \nu$ on $X \times Y$ has the following properties for A, B, C Borel sets:-*

(i) $\mu \times \nu(A \times B) = \mu(A) \times \nu(B)$;
(ii) $\mu \times \nu(C) = \int_X (x \mapsto \nu(C_x))d\mu$;
(iii) $\mu \times \nu(C) = \int_Y (y \mapsto \mu(C^y))d\nu$.

Exercise Show that the conclusions (ii) and (iii) hold also if we simply assume $C \subset X \times Y$ is measurable.

Finally! We come to Fubini's theorem:

Theorem 13.7 *(Fubini) Let (X, μ) and (Y, ν) be standard Borel probability spaces and let*

$$f : X \times Y \to \mathbb{R}$$

be a measurable non-negative function. Then

$$\int_{X\times Y} fd(\mu \times \nu) = \int_X (x \mapsto (\int_Y y \mapsto f(x,y)d\nu)d\mu = \int_Y (y \mapsto (\int_X x \mapsto f(x,y)d\mu)d\nu.$$

Proof We will traverse through a sequence of special cases, finally obtaining the result for general f.

First of all if f has the form

$$a \cdot \chi_C,$$

for $a \in \mathbb{R}$, χ_C the characteristic function of some measurable set, the result is presented at 13.6. Similarly in the case that f is a finite sum of such functions, a simple function, we obtain the result by the additivity of integration. Finally in the general case we can obtain a sequence of simple functions, (f_n), with

$$0 \leq f_i(x) \leq f_{i+1}(x) \leq f(x)$$

and $f_n(x) \to f(x)$ at every x. Now the result follows from 5.3. □

I have been rather fussy in the statement of Fubini above, almost to the point of being obsessively explicit. Usually one would just put this as

$$\int_{X\times Y} fd(\mu \times \nu) = \int_X \int_Y fd\nu d\mu = \int_Y \int_X fd\mu d\nu.$$

We are not quite ready to state the measure disintegration theorem, but it is possible to give a sense of its content. Suppose we have measure on $X \times Y$. Call it m. Suppose μ is a measure on X. Now there is no guarantee that there will be a ν on Y for which $m = \mu \times \nu$. Just doesn't necessarily hold. Even under the additional assumption of

$$m(B \times Y) = \nu(B)$$

for all Borel $B \subset X$.

However something else is true. We can find a range of measures, ν_x on Y, for each $x \in X$ such that m equals the integral, as it were, along μ of the various ν_x's. More precisely, for $A \subset X \times Y$,

$$m(A) = \int_X (x \mapsto \nu_x(A_x))d\mu.$$

There are some subtleties here being swept under the rug by our notation. To even make sense of this we need to know that $x \mapsto \nu(A_x)$ will always be measurable, and this in turn needs us to have some idea of what it would mean for the function $x \mapsto \nu_x$ to be Borel or measurable. It won't be until we see the Riesz representation theorem that this can be made precise.

A slightly more sophisticated version of the lemma applies to measure preserving functions. Suppose $(X, \mu), (Z, \lambda)$ are standard Borel probability spaces and

$$f : X \to Z$$

is measure preserving function. As a matter of notation, let $Y_z = \{x \in X : f(x) = z\}$ for any $z \in Z$. Then our more sophisticated theorem states this: We can find a (suitably measurable) assignment $z \mapsto \mu_z$, where each μ_z is a measure on Y_z, such that

$$\mu(A) = \int_Z (z \mapsto \mu_z(A \cap Y_z))d\lambda.$$

14 Measure disintegration

One of the most important consequences of 11.5 is ontological as much as mathematical. We have a way of thinking of the probability measures on a standard Borel space as a standard Borel space in its own right. Given X Polish we can find a compact metric space K which is Borel isomorphic to X, allowing us to identify the Borel probability measures on X with $P(K)$. From the exercise following 11.6, $P(K)$ can itself be viewed as a compact metric space, and in the natural indentification, $P(X)$, the Borel probability measures on X, can be viewed as a standard Borel space.

Lemma 14.1 *Let X be a Polish space. Then the Borel sets are those appearing in the smallest collection containing the open sets and closed under complements and countable* disjoint *unions.*

Proof Let Σ be the smallest collection containing the open sets and closed under the operations of complements and countable disjoint union. It suffices to show that this collection of sets forms an algebra.

For any $A \in \Sigma$, let $\Sigma(A) = \{B \subset X : B \cap A \in \Sigma\}$.

Claim: $\Sigma(A)$ is closed under complements and countable disjoint unions.

Proof of Claim: The main issue is complementation. But $A \setminus B$ equals $X \setminus (X \setminus A \cup (A \cap B))$, and hence will be in Σ assume $B \cap A$ is. (Claim$\square$)

If U is open, it then follows that $\Sigma(U)$ includes the open sets, and hence includes Σ. In particular, we obtain for any $B \in \Sigma$ that $U \cap B$ is in Σ.

Turning things around from the point of view of B and considering the open sets, we obtain that for any $B \in \Sigma$ we have the open sets included in $\Sigma(B)$, and hence Σ will be included in $\Sigma(B)$ – which is to say, that for any $A \in \Sigma$,

$$A \cap B \in \Sigma,$$

which is all we need to establish Σ as an algebra. $\square$

Theorem 14.2 *Let K be a compact metric space. Then for any bounded Borel function, the resulting map*

$$P(K) \to \mathbb{R}$$

$$\mu \mapsto \int f(x) d\mu(x)$$

is Borel.

Proof It suffices to see that for any Borel set $B \subset K$

$$P(K) \to \mathbb{R}$$

$$\mu \mapsto \mu(B)$$

is Borel.

For any open set $U \subset K$ we can find a sequence of continuous functions $(f_n)_{n \in \mathbb{N}}$ with $0 \leq f_n \leq 1$, $f_n \leq f_{n+1}$, $f_n \prec U$, and

$$U = \bigcup_{n \in \mathbb{N}} \{x : f_n(x) = 1\}.$$

It then follows that $\mu(U) = \lim_{n \to \infty} \int f_n d\mu$. Since

$$\mu \mapsto \int f d\mu$$

is continuous, and certainly Borel, for any $f \in C(K)$, we indeed have $\mu \mapsto \mu(U)$ as a Borel function.

For the remainder of the Borel sets, it suffices in light of the last lemma to verify that the collection of Borel sets for which $\mu \mapsto \mu(B)$ is closed under complements and disjoint unions.

For complements, $\mu(K \setminus B) = 1 - \mu(B)$. And for disjoint unions, if $(B_n)_{n\in\mathbb{N}}$ are disjoint, then

$$\mu(\bigcup_{n\in\mathbb{N}} B_n) = \sum_{n\in\mathbb{N}} \mu(B_n) = \lim_{N\to\infty} \sum_{n\leq N} \mu(B_n).$$

□

Recall that a *Borel isomorphism* $f : X \to Y$ between Polish spaces X and Y is a bijection which exactly preserves Borel structure: $B \subset X$ is Borel if and only if $f[B] = \{f(x) : x \in B\}$ is Borel.

Corollary 14.3 *Let K_1, K_2 be compact metric spaces. Let*

$$f : K_1 \to K_2$$

be a Borel isomorphism. Then f induces a Borel isomorphism

$$f^* : P(K_1) \to P(K_2)$$

via

$$(f^*(\mu))(B) = \mu(f^{-1}[B]).$$

Proof Let $(f_i)_{i\in\mathbb{N}}$ be a countable dense subset of $C(K_2, [-1, 1])$, the elements of $C(K_2)$ with sup norm at most one. Any $\nu \in P(K_2)$ is canonically determined by its behavior on these elements, as discussed in the exercises following the proof of 11.5, and so we only need to show that

$$\hat{f} : P(K_1) \to \prod_{i\in\mathbb{N}} [-1, , 1]$$

given by

$$(\hat{f}(\mu))(n) = (f^*(\mu))(f_n)$$

is Borel. However if we let $h_n = f_n \circ f$ then

$$\mu \mapsto \int h_n d\mu = \int f_n df^*(\mu)$$

is Borel by the theorem above. □

Thus we have for any standard Borel space X a canonical standard Borel structure on $P(X)$, the collection of all Borel probability measures on X: Apply 7.1 to find some compact metric K which admits a Borel bijection with X, and then take the Borel structure on $P(K)$. The key point here is given by the above lemma: The resulting Borel structure on $P(X)$ is unaffected by the circumstances surrounding our choice of K.

Theorem 14.4 *Let (X, μ) be a standard Borel probability space. Let*

$$f : X \to Y$$

be Borel with $\nu = f^*[\mu]$ in $P(Y)$ defined by

$$\nu(B) = \mu(f^{-1}[B]).$$

Then there is a Borel function

$$Y \to P(X)$$

$$y \mapsto \mu_y$$

with

$$\mu(A) = \int \mu_y(A) d\nu(y)$$

for any Borel $A \subset X$.

Proof The proof is comparatively trivial in the case that X is countable, so assume instead it is uncountable, and then by 7.1 we may assume X equals Cantor space, $2^{\mathbb{N}}$. We let $(C_n)_{n\in\mathbb{N}}$ enumerate the clopen subset of $2^{\mathbb{N}}$. For each $C \in \{C_n : n \in \mathbb{N}\}$ we let

$$\nu_C(B) = \nu(C \cap f^{-1}(B),$$

to obtain a positive measure on Y. Note that

$$\nu_{C\cup C'}(B) = \nu_C(B) + \nu_{C'}(B)$$

for C, C' disjoint.

Each such ν_C is clearly absolutely continuous with respect to ν, and so we can apply 6.4 to find Borel functions $(f_C)_{C \text{ clopen}}$ with

$$\int_B f_C(y) d\nu(y) = \nu_C(B).^3$$

Claim: For $(A_i)_{i\leq k}$ disjoint clopen sets

$$f_{\bigcup_i A_i} = \sum_i f_{A_i}$$

almost everywhere.

Proof of Claim: For any Borel B we have

$$\int_B f_{\bigcup_i A_i}(y) d\nu(y) = \nu_{\bigcup_i A_i}(B)$$

$$= \sum_i \nu_{A_i}(B) = \sum \int_B f_{A_i}(y) d\nu$$

$$= \int_B \sum f_{A_i}(y) d\nu.$$

Thus it is impossible to find a non-null Borel set B on which either $f_{\bigcup_i A_i} < \sum_i f_{A_i}$ or $f_{\bigcup_i A_i} > \sum_i f_{A_i}$, as required. (Claim□)

Thus for a conull set of y we can define

$$\mu_y^* : \{C_n : n \in \mathbb{N}\} \to [0,1]$$

$$C_n \mapsto f_{C_n}(y)$$

[3] You may object. Literally as stated at 6.4, we only obtain measurable functions. But every measurable function is Borel on a conull Borel set and we will suffer no harm if we simply adjust them off of the countable union of all the conull subsets involved.

which will be finitely additive in the natural sense. This uniquely extends to a linear function

$$\mu_y : C(2^{\mathbb{N}}) \to \mathbb{R}$$

since every continuous function on Cantor space can be approximated in the sup norm by a continuous function with only finitely many values.

I will leave as an exercise the tedious computations and untangling of the definitions necessary to verify

$$y \mapsto \mu_y$$

Borel.[4]

Then we have that

$$C \mapsto \int \mu_y(C) d\nu(y)$$

agrees with μ on the clopen sets, and using that they are both measures we have for every Borel A

$$\mu(A) = \int \mu_y(A) d\nu(y).$$

□

[4] But please do see me if you do not feel completely clear about what is going on.

15 Haar measure

Definition A topological space G equipped with the group operations of multiplication

$$G \times G \to G$$

$$(g, h) \mapsto gh$$

and inverse

$$G \to G$$

$$g \mapsto g^{-1}$$

is said to be a *topological group* if both these operations are continuous. We say that it is a *lcsc* group if in addition as a topological space it is Hausdorff, separable, and locally compact.

It follows from the Urysohn metrizability criteria, as can be found in say [4], that any lcsc group admits a compatible metric, and from there the locally compactness gives that it admits a complete compatible metric.

Definition For G a lcsc, σ-finite measure μ defined on its Borel subsets is said to be a *left Haar measure* if it is not constantly zero and for any Borel $B \subset G$ and $g \in G$ we have

$$\mu(B) = \mu(g \cdot B),$$

where here $g \cdot B =_{\text{df}} \{gh : h \in B\}$ is the left translation of B by g. Similarly a non-trivial σ-finite measure on the Borel sets of G is said to be a *right Haar measure* if it is invariant under right translation.

We will show the existence of Haar measures for lcsc groups. The proof will be organized around the notion of *content* from §11. Recall that $\mathcal{K}(X)$ denotes the compact subsets of a topological space X.

Theorem 15.1 *Let G be a lcsc group. Then there is a content*

$$\rho : \mathcal{K}(G) \to \mathbb{R}^{\geq 0}$$

with the properties that $\rho(K) \neq 0$ for some $K \in \mathcal{K}(G)$ and that for any $g \in G, K \in \mathcal{K}(G)$

$$\rho(K) = \rho(g \cdot K).$$

Proof Notation: For $B_1, B_2 \subset G$ we let $B_1 : B_2$ be the smallest natural number $n \in \mathbb{N}$, if it exists, such that there are $g_1, g_2, ..., g_n \in G$ with

$$B_1 \subset \bigcup_{i \leq n} g_i \cdot B_2.$$

If no such finite n exists we let $B_1 : B_2 = \infty$.

Note here that for any such B_1, B_2 and $g \in G$, the structure of the definition gives

$$B_1 : B_2 = g \cdot B_1 : B_2.$$

Claim: If O is an open non-empty subset of G and K is a compact subset of G, then

$$K : O < \infty.$$

Proof of Claim: We may assume that O contains the identity. Then at each $g \in K$ we have $g \in g \cdot O$. Thus

$$K \subset \bigcup_{g \in K} g \cdot O,$$

and the claim follows by the compactness of K. (Claim$\Box$)

Let W be an open subsets of 1_G, the identity for G, whose closure is compact. Let $A = \overline{W}$ be its closure.

Notation: For $O \subset G$ an open neighborhood of the identity we let

$$\lambda_O : \mathcal{K}(G) \to \mathbb{R}^{\geq 0}$$

$$K \mapsto \frac{K : O}{A : O}.$$

We then let $\mathcal{D}(O) = \{\lambda_U : U \subset O \text{ open}, 1_G \in U\}$ for any open non-empty set O.

Note that any non-empty open O has $\lambda_O(A) = 1$ and $\lambda_O(K) \leq K : A$ for any $K \in \mathcal{K}(G)$. We then let

$$C = \prod_{K \in \mathcal{K}(G)} [0, K : A],$$

equipped with the product topology derived by viewing each closed interval $[0, K : A]$ as a topological space in the usual euclidean topology. Thus every λ_O can be viewed as an element of C and every $\mathcal{D}(O) \subset C$.

Claim:

$$\bigcap \{\overline{\mathcal{D}(O)} : O \subset G, O \text{open}, 1_G \in O\}$$

is non-empty.

Proof of Claim: Here $\overline{\mathcal{D}(O)}$ refers to the closure in C.

By Tychonov's theorem we know that C is compact, and thus it suffices to show that for any finite collection F of open neighborhoods of the identity we have

$$\bigcap_{O \in F} \overline{\mathcal{D}(O)} \neq \emptyset.$$

However this follows since $\bigcap_{O \in F} \mathcal{D}(O) \supset \mathcal{D}(\bigcap_F O)$. (Claim$\Box$)

Claim: If K_1, K_2 are disjoint open sets then there exists an open neighborhood of the identity O such that for all $\rho \in \mathcal{D}(O)$

$$\rho(K_1 \cup K_2) = \rho(K_1) + \rho(K_2).$$

Proof of Claim:

Subclaim: There exists an open neighborhood O of 1_G such that

$$O^{-1} O K_1 \cap K_2 = \emptyset.$$

Proof of subclaim: At each $g \in K_1$ we can by the continuity of the group operations choose an open U_g containing 1_G and open W_g containing g such that

$$U_g W_g \cap K_2 = \emptyset.$$

(Here $U_g W_g =_{\mathrm{df}} \{h_1 h_2 : h_1 \in U_g, h_2 \in W_g\}$.) At each $g \in K_1$ fix an open neighborhood O_g of 1_G with $O_g^{-1} O_g \subset U_g$. Now compactness enables us to choose a finite $F \subset K_1$ such that

$$K_1 \subset \bigcup_{g \in F} W_g.$$

Then we take $O = \bigcap_{g \in F} O_g$. (□Proof of subclaim)

Now for any open non-empty $U \subset O$ we have that if $g \in G$ and

$$h_1, h_2 \in g \cdot U$$

then $h_1^{-1} h_2 \in U^{-1} U \subset O^{-1} O$. Hence, $g \cdot U$ cannot simultaneously have non-empty intersection with both K_1 and K_2. It follows then from the definition of λ_U that

$$\lambda_U(K_1 \cup K_2) = \lambda_U(K_1) + \lambda(K_2).$$

(Claim□)

Now consider the following conditions for $K_1, K_2 \in \mathcal{K}(X)$ and $g \in G$:

1. $\rho(K_1) + \rho(K_2) = \rho(K_1 \cup K_2)$ when $K_1 \cap K_2 = \emptyset$;
2. $\rho(K_1) \leq \rho(K_2)$ when $K_1 \subset K_2$;
3. $\rho(K_1) = \rho(g \cdot K_1)$.;
4. $\rho(A) = 1$.

These are all closed conditions on the functions $\rho \in C$. The second, third, and fourth are satisfied by any λ_U for U open. The first is satisfied by any λ_U for U a subset of a sufficiently small open set. Thus any

$$\rho \in \bigcap \{\overline{\mathcal{D}(O)} : O \subset G, O \text{open}\}$$

is as required to complete the proof of the theorem. □

Corollary 15.2 *Any lcsc topological group admits a left Haar measure and a right Haar measure.*

Proof It suffices to do the case of left Haar measure. Applying 15.1 we obtain a left invariant non-trivial content. Then applying 11.3 we obtain a measure which is induced from the outer measure derived from the content, which then by the structure of the definitions will be left invariant. □

Examples 1. For $(\mathbb{R}, +)$, the additive group of the reals, a Haar measure is given by Lebesgue measure. Here there is no need to distinguish left Haar measure from right Haar measure, since the notions coincide in virtue of the group being abelian.

2. In higher dimensions the higher dimensional Lebesgue measures again provide a Haar measure. In $(\mathbb{R}^n, +)$ the n^{th} dimensional Lebesgue measure m_n serves as a Haar measure.

3. For $(\mathbb{R}^{>0}, \cdot)$. the multiplicative group of the positive reals, the expression is more complicated. One way to describe a Haar measure is by

$$\mu(A) = \int_A \frac{1}{t} dm(t),$$

where m is the Lebesgue measure.

4. For $GL_n(\mathbb{R})$, the invertible n by n matrices with real valued entries, one obtains a Haar measure by

$$\mu(A) = \int_A det(M) dm_{n^2}(M),$$

where m_{n^2} is the measure on $M_{n,n}(\mathbb{R})$, the full set of all n by n matrices with real valued entries, obtained under the natural isomorphism between $M_{n,n}(\mathbb{R})$ and $\mathbb{R}^{n^2}$ which we can equip with Lebesgue measure.

5. In the case of discrete groups, a Haar measure is given simply by the counting measure. So for Γ a discrete countable group, we let
$$\mu(A) = |A|,$$
the cardinality of A, for any $A \subset \Gamma$.

6. A version of this holds for products of finite groups. So given $(\Gamma_n)_{n\in\mathbb{N}}$ a sequence of finite groups, we can view them all as discrete groups, and then let
$$\Gamma = \prod_{n\in\mathbb{N}} \Gamma_n$$
be the resulting group equipped with the product topology. Then given a basic open set,
$$U = \{f \in \Gamma : f(1) \in A_1, f(2) \in A_2, ..., f(n) \in A_n\},$$
for $A_1, ...A_n$ subsets of $\Gamma_1, ...\Gamma_n$, we let
$$\mu(U) = \frac{|A_1| \cdot |A_2|.... \cdot |A_n|}{|\Gamma_1| \cdot |\Gamma_2|... \cdot |\Gamma_n|}.$$
Then it follows from the extension theorems established in the earliest chapters that μ extends uniquely to a Borel probability measure on Γ, and it is relative straightforward to see that this measure must be both left and right invariant.

Exercise Let μ be a left Haar measure on a lcsc group Γ. Show that if K is compact then $\mu(K)$ is finite.

There are further theorems regarding uniqueness.

Theorem 15.3 *Let Γ be a lcsc group and U a non-empty open subset of Γ with compact closure. Then for any $\alpha > 0$ there is a unique left Haar measure with $\mu(U) = \alpha$, and similarly there is a unique right Haar measure which assigns the value α to U.*

Theorem 15.4 *Let Γ be a compact lcsc group. Then any left Haar measure is a right Haar measure.*

Definition A lcsc group G is *unimodular* if it has a left Haar measure which is also a right Haar measure.

So the last theorem states that the compact groups are unimodular, and clearly the abelian groups are unimodular. However there are many examples of lcsc groups which are not unimodular, and one can even find nilpotent lcsc groups which are not unimodular.

Exercise Let Γ be the group of 2 by 2 matrices of the form
$$\begin{bmatrix} x & y \\ 0 & x^{-1} \end{bmatrix}$$
where $x, y \in \mathbb{R}$, $x \neq 0$. Show that Γ is not unimodular.

16 Ergodic theory

16.1 Examples and the notion of recurrence

Definition Let (X, μ) be a standard Borel probability space and let $T : X \to X$ be a measurable transformation. We say that T is *measure preserving* if $\mu(T^{-1}[A]) = \mu(A)$ for every measurable $A \subset X$. A measurable set $A \subset X$ is said to be T*-invariant* if

$$T^{-1}[A] = A.$$

We then say that the system (X, μ, T) is *ergodic* if every invariant measurable set is either null or conull.

Examples 1. Let $X = 2^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$, equipped with the product topology and the product measure: Given a cylinder set $A = \{f : f(1) = S_1, f(2) = S_2, ...f(N) = S_N\}$ we let $\mu(A) = 2^{-N}$; there is a unique Borel measure extending μ to the collection of Borel sets. Then let T be the one sided shift:

$$(T(f))(n) = f(n+1).$$

This is almost immediately seen to be measure preserving. For ergodicity, we let $A \subset X$ be a measurable set of measure between ϵ and $1-\epsilon$, some $\epsilon > 0$. We can find finite sequences of cylinder sets, $A_1, ..., A_N$ such that

$$\mu(A \Delta (\bigcup_{i \leq N} A_i)) < \epsilon^2$$

At some large n we get

$$T^{-n}(\bigcup_{i \leq N} A_i)$$

independent, in the sense of measure theory, from $\bigcup_{i \leq N} A_i$. Thus

$$\mu(T^{-n}(\bigcup_{i \leq N} A_i) \setminus \bigcup_{i \leq N} A_i) \geq \epsilon^2,$$

$$\therefore \mu(T^{-n}(A) \setminus A) > 0.$$

2. Let $X = \mathbb{R}/\mathbb{Z}$, with the quotient topology and Lebesgue measure, λ. Let

$$T : x \mapsto x + \sqrt{2} \bmod 1.$$

It is easily seen that T acts by isometries. Since $\sqrt{2}$ is irrational, the *orbit* of a point x,

$$\{T^{\ell}(x) : \ell \in \mathbb{Z}\},$$

is always infinite.

There are various proofs of ergodicity, though none of them are immediately obvious.

I am going to give one argument using Hilbert space theory. Let $\mathcal{H}$ be the Hilbert space of all measurable, square integrable functions from (X, λ) to $\mathbb{C}$. For each $\ell \in \mathbb{Z}$ let

$$\pi_\ell : x \mapsto e^{2\ell\pi i x} = \cos(2\ell\pi x) + i\sin(2\ell\pi x).$$

It is a routine calculation to see that $\langle \pi_\ell, \pi_k \rangle = 0$ for $\ell \neq k$, and so we certainly have an orthonormal set. In fact this forms an orthonormal basis.[5] One way to see this is to apply Stone-Weierstrass at 12.5 to conclude

[5] Here and elsewhere I am blurring over the distinction between a measurable function and its *equivalence class* in the corresponding Hilbert space $L^2(X, \mu)$.

that the finite linear combinations of these functions are dense in $C(X, \mathbb{C})$, and then it is standard that $C(X, \mathbb{C})$ is dense in $\mathcal{H}$.

So now suppose $A \subset X$ is a measurable, invariant set. Let χ_A be the characteristic function of this set. $\chi_A \in \mathcal{H}$ and so we can find coefficients $(c_\ell)_{\ell \in \mathbb{Z}}$ such that

$$\chi_A(x) = \sum_{\ell \in Z} c_\ell \pi_\ell(x)$$

almost everywhere.

$$\chi_A \circ T = T$$

by invariance, whilst

$$\pi_\ell \circ T = e^{2\ell\pi i\sqrt{2}} \pi_\ell,$$

which unwinds to give us

$$\sum_{\ell \in Z} c_\ell \pi_\ell(x) = \sum_{\ell \in Z} e^{2\ell\pi i\sqrt{2}} c_\ell \pi_\ell(x)$$

almost everywhere. Since $\{\pi_\ell : \ell \in \mathbb{Z}\}$ is an orthonormal basis, we obtain $c_\ell = e^{2\ell\pi i\sqrt{2}} c_\ell$ at every ℓ, which entails $c_\ell = 0$ all $\ell \neq 0$, yielding χ_A constant almost everywhere. Just as required.

Notation We say that T is an *m.p.t.* on (X, u) if it is a measurable, measure preserving function from X to X.

Lemma 16.1 *(Poincare recurrence lemma) Let (X, μ) be a standard Borel probability space. Let T be an m.p.t. on (X, μ). Let $A \subset X$ be measurable and non-null.*

Then for almost every $x \in A$ there exists some $n > 0$ with

$$T^n(x) \in A.$$

Proof Suppose otherwise, and let A_n be the set of x for which

$$T^n(x) \in A$$

but at all $k > n$

$$T^k(x) \notin A.$$

Note that $A_0 = \{x \in A : \forall n > 0 T^n(x) \notin A\}$, which we are assuming to be positive.

$$A_n = T^{-n}[A] \cap \bigcap_{k>n} (X \setminus T^{-k}[A]),$$

and so is certainly measurable.

$$T^{-1}[A_n] = A_{n+1},$$

so they all have the same measure as A_0, and for $k \neq n$ we have $A_n \cap A_k$ empty.

Thus $(A_n)_{n \in \mathbb{N}}$ is an infinite sequence of disjoint measurable sets, all with the same non-zero measure, contradicting finiteness of μ. □

There is a peculiar consequence of this simple lemma. Suppose I start with a two chambered tank of gas. I pump all the air out of one of the chambers, and then remove the partition between the two chambers. Intuitively we expect the air inside the chamber to rapidly spread out equally between the two chambers. Indeed the second law of thermodynamics essentially predicts a mixing up and a diffusion of the air.

However, at the most basic level, we have a dynamical system. The particles are moving around through time, and although the event of all the air being in one of the two chambers and none in the other is fantastically implausible and unlikely, it is possible. The Poincare recurrence lemma predicts that with enough time this event must reoccur.

Take another model which is a bit more mathematical. I have two urns, one blue, one red. I start with 100 ping pong balls all in the red urn. Every five seconds we toss a coin for each ball. If the coin toss for that ball comes up heads, then we move it to the other urn. If it was in a blue urn and we tossed heads, it switches to the red urn. If in the red urn and we tossed for it a head, it moves to the blue urn. If we toss a tail, then it stays foot.

Intuitively we expect a mixing. After the first five seconds, we would expect about half the balls to immediately move across to the blue urn having received a head in their toss. Perhaps after a while we would imagine relatively stable populations, clustered around 50 balls in each. Again this is exactly the prediction of the second law of thermodynamics.

But, but, but, Poincare tells us something else. If we wait long enough, the event of all the balls being in the red urn and none in the blue is not only possible but *inevitable*. If we wait long enough, it will happen – again and again and again.

There is an extended discussion of this example in §2.3 of [8]. He points out that the expected return, the period in purgatory while we wait for the red to fill completely and the blue to clear out, is far, far longer than the likely age of the universe. The second law of thermodynamics may be literally false, but as a rule of thumb it holds fairly true.

16.2 The ergodic theorem and Hilbert space techniques

On to further topics.

I am going to base the rest of the discussion around Hilbert space techniques. Ultimately I want to work towards the investigation of general ergodic actions of countable groups and make the connection between this kind of ergodic theory and certain results in *percolation theory*.

To speed up the journey ahead, I am going to confine the discussion to *invertible* m.p.t's. Many of the results below hold in a more general setting and that more general setting is considered important to ergodic theorists, but we will be working under simplifying assumptions. The goal is to see the main ideas simply rather than the best possible results in all their glory and complexity.

A quick review of some basic ideas from Hilbert space. Probably you know these already but you can find them discussed in any reasonably advanced book on linear algebra. Much of it is in §9.2 of [7]. The first chapter of [3] gives a pretty thorough account.

From here on I will be taking the Hilbert spaces over $\mathbb{C}$ rather than $\mathbb{R}$. Most of the time this makes only a nominal difference, but later on it will be important to have things framed in this way: By working over $\mathbb{C}$ we ensure full access to all potential eigenvalues.

Definition Let $\mathcal{H}$ be a Hilbert space and $\mathcal{H}_0$ a closed subspace. We then let $\mathcal{H}_0^\perp$ be the collection of $f \in \mathcal{H}$ for which

$$\forall g \in \mathcal{H}_0(\langle f, g\rangle = 0).$$

It is then a standard fact that every $f \in \mathcal{H}$ can be resolved uniquely into the form

$$f = f_0 + f_1$$

where $f_0 \in \mathcal{H}_0$ and $f_1 \in \mathcal{H}_0^\perp$. We can then define

$$P : \mathcal{H} \to \mathcal{H}_0$$

by letting $Pf = f_0$, where f_0 is as above. P is called the *orthogonal projection* to $\mathcal{H}_0$.

Lemma 16.2 *Let $\mathcal{H}$ be a Hilbert space and $\mathcal{H}_0$ a closed subspace Then:*
(i) the orthogonal projection $P : \mathcal{H} \to \mathcal{H}_0$ is a linear contraction;
(ii) $(\mathcal{H}_0^\perp)^\perp = \mathcal{H}_0$.

Recall that in a Hilbert space we can define the Hilbert space norm *from* the inner product:

$$\|f\| = (\langle f, f\rangle)^{\frac{1}{2}}.$$

Lemma 16.3 *(Cauchy-Schwarz) For $\mathcal{H}$ a Hilbert space and $f, g \in \mathcal{H}$*

$$|\langle f, g\rangle| \leq \|f\|\|g\|.$$

Moreover, equality only occurs when f and g are scalar multiples of one another.

Definition For $\mathcal{H}$ a Hilbert space and $U : \mathcal{H} \to \mathcal{H}$ bounded linear operator, we define

$$U^* : \mathcal{H} \to \mathcal{H}$$

by the formula

$$\langle U^* f, g\rangle = \langle f, Ug\rangle.$$

(It is a standard fact that this U^* *is* well defined and is itself a bounded linear operator.)

We say that U is *unitary* if it is onto and for all $f, g \in \mathcal{H}$

$$\langle Uf, Ug\rangle = \langle f, g\rangle.$$

Note then that U will also be one to one ($\|Uf\| = \|f\|$ all $f \in \mathcal{H}$) and hence invertible.

Lemma 16.4 *If $U : \mathcal{H} \to \mathcal{H}$ is unitary, then $U^* = U^{-1}$.*

Definition Let (X, μ) be a standard Borel probability space and $T : X \to X$ an invertible m.p.t. We then define

$$U_T : L^2(X, \mu) \to L^2(X, \mu)$$

by

$$U_T(f)(x) = f(T(x)).$$

Lemma 16.5 *For T as above, U_T is a unitary operator.*

So much for the basics. Now for some true grit.

Lemma 16.6 *Let $\mathcal{H}$ be a Hilbert space and*

$$U : \mathcal{H} \to \mathcal{H}$$

unitary. Let $\mathcal{H}_0 = \{f \in \mathcal{H} : Uf = f\}$, the closed subspace of U-invariant vectors. Let

$$P : \mathcal{H} \to \mathcal{H}_0$$

be the orthogonal projection.

Then for each $f \in \mathcal{H}$

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k f \to Pf.$$

Proof Let
$$\mathcal{H}_1 = \{g - Ug : g \in \mathcal{H}\}.$$
It is routinely verified that this is a closed subspace of $\mathcal{H}$.

Claim: $\mathcal{H}_1^\perp = \mathcal{H}_0$.

Proof of Claim: First consider $f \in \mathcal{H}_1^\perp$. We have
$$\langle f, f - Uf\rangle = 0$$
$$\therefore \langle f, f\rangle = \langle f, Uf\rangle.$$
Since $\|Uf\| = \|f\|$ it follows from Cauchy-Schwarz that f and Uf are scalar multiples and from there we quickly obtain their equality.

Conversely, suppose $f \in \mathcal{H}_0$. Using U unitary we see for $g \in \mathcal{H}$
$$\langle f, g - Ug\rangle = \langle f, g\rangle - \langle f, Ug\rangle$$
$$= \langle f, g\rangle - \langle f, Ug\rangle = \langle f, g\rangle - \langle U^* f, g\rangle$$
$$= \langle f, g\rangle - \langle U^{-1} f, g\rangle = 0$$
since $f = Uf = U^{-1}f$. (Claim□)

Note that this claim does not yet license us to draw the conclusion that $\mathcal{H}_1 = \mathcal{H}_0^\perp$. That would only be true if we knew $\mathcal{H}_1$ to be closed. *Instead* we are in a position to conclude only that $\overline{\mathcal{H}_1} = \mathcal{H}_0^\perp$, where $\overline{\mathcal{H}_1}$ is the closure of $\mathcal{H}_1$ in the topology provided by the Hilbert space norm.

Claim: For $f \in \overline{\mathcal{H}_1}$,
$$\frac{1}{n}\sum_{k=0}^{n-1} U^k f \to 0.$$

Proof of Claim: Fix $\epsilon > 0$. Choose $\bar{f} \in \mathcal{H}_1$, $\bar{f} = g - Ug$, with $\|f - \bar{f}\| < \epsilon/2$. Then find some N with
$$\frac{2\|g\|}{N} < \frac{\epsilon}{2}$$
and hence
$$\|\frac{1}{n}\sum_{k=0}^{n-1} U^k \bar{f}\| = \|\frac{1}{n}(g - U^n g)\| \leq \frac{1}{n}(\|g\| + \|U^n g\|) =\leq \frac{1}{n}(\|g\| + \|g\|) < \frac{\epsilon}{2}.$$
all $n > N$. Then
$$\|\frac{1}{n}\sum_{k=0}^{n-1} U^k f\| \leq \|\frac{1}{n}\sum_{k=0}^{n-1} U^k (f - \bar{f})\| + \|\frac{1}{n}(g - U^n g)\|$$
$$\leq \frac{1}{n}\sum_{k=0}^{n-1} \|f - \bar{f}\| + \frac{\epsilon}{2} < \epsilon$$
for all $n > N$. (Claim□)

For a general $f \in \mathcal{H}$, write
$$f = f_0 + f_1,$$

where $f_0 = Pf \in \mathcal{H}_0$ and $f_1 \in \overline{\mathcal{H}_1} = \mathcal{H}_0^\perp$.

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k f = \frac{1}{n}\sum_{k=0}^{n-1} U^k f_0 + \frac{1}{n}\sum_{k=0}^{n-1} U^k f_1$$

$$= f_0 + \frac{1}{n}\sum_{k=0}^{n-1} U^k f_1 \to f_0.$$

□

Definition A n m.p.t $T : X \to X$ is said to be *invertible* if it is one-to-one, onto, and its inverse is also an m.p.t.

It actually follows from Lusin Novikov at 8.10 that a one-to-one m.p.t on a standard Borel probability space will necessarily be invertible. I will not use this particular fact.

Theorem 16.7 *(The von Neumann ergodic theorem) Let (X, μ) be a standard Borel probability space and $T : X \to X$ an invertible m.p.t. Let $f \in L^2(X, \mu)$. Then there is $\bar{f} \in L^2(X, \mu)$ with*

$$\lim_{n\to\infty} \| \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k - \bar{f} \|_2 \to 0.$$

Proof Let U_T be the induced unitary operator on $L^2(X, \mu)$:

$$(U_T(g))(x) = g(T(x)).$$

Then the last theorem gives us some $\bar{f} = Pf$ with

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k f \to Pf$$

$$\therefore \frac{1}{n}\sum_{k=0}^{n-1} U^k f - Pf \to 0$$

in $L^2(X, \mu)$. □

There are many minor variations of this result. It is not relevant for us to trek through them all in detail. Let me at least mention passingly that the assumption of T being invertible can be dropped – though in that case the corresponding U_T might not be unitary and we need to prove a version of 16.6 suitable for U being a linear contraction. In the case that T *is* invertible we can actually obtain convergence of the partial sums

$$\frac{1}{2n+1}\sum_{k=-n}^{k=n} f \circ T^k$$

in $L^2(X, \mu)$.

Lemma 16.8 *Let (X, μ) be a standard Borel probability space and $T : X \to X$ an invertible, ergodic, m.p.t. Let $\bar{f} \in L^2(X, \mu)$ be T-invariant. Then $\bar{f}$ is constant almost everywhere.*

Proof It suffices to show that for every Borel set, $\bar{f}^{-1}[B]$ is either null or conull. But this is a direct consequence of ergodicity. $\square$

Now we have a crucial consequence of the von Neumann ergodic theorem.

Corollary 16.9 *Let (X, μ) be a standard Borel probability space and $T : X \to X$ an invertible. ergodic, m.p.t. Let $f \in L^2(X, \mu)$. Let $\alpha = \int f d\mu$. Then*

$$\int |\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k - \alpha| d\mu \to 0.$$

Proof First let us take the $\bar{f}$ from the von Neumann ergodic theorem. Let

$$f_n = \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k.$$

Claim: $\bar{f}$ is T-invariant.

Proof of Claim: First note that

$$\|f_n - f_n \circ T\|_2 = (\langle \frac{1}{n}(f - f \circ T^{k+1}), \frac{1}{n}(f - f \circ T^{k+1})\rangle)^{\frac{1}{2}}$$

$$\leq \frac{\sqrt{2}}{n}\|f\|_2 \to 0.$$

Thus since $f_n \to \bar{f}$ and U_T is an isometry,

$$\bar{f} = U_T \bar{f}.$$

(Claim$\square$)

Claim: $\int (f_n - \bar{f}) d\mu \to 0$.

Proof of Claim: For any $g \in L^2(X, \mu)$ we have

$$\langle f_n - f, g\rangle \to 0$$

by Cauchy-Schwarz. In particular,

$$\langle f_n - \bar{f}, 1\rangle \to 0,$$

which is exactly the same as saying $\int (f_n - \bar{f}) d\mu \to 0$. (Claim$\square$)

Now each $\int f_n d\mu = \int f d\mu$, and the constant function $\bar{f}$ has no choice but to be equal to $\int f d\mu$ a.e. $\square$

Here there is a famous slogan: If T is ergodic, then "a.e. the time mean of f equals the space mean of f." The time mean here refers to starting at a point x and taking the averages

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)),$$

sampling through iterations under the map. The space mean of course refers to the integral of f. This identity is one of the most fundamental in ergodic theory, with perennial applications. Before presenting the proof of this result we need a curious technical lemma.

Lemma 16.10 *Let T be an m.p.t. on a standard Borel probability space (X, μ). Let $f \in L^1(X, \mu)$ and let B be the set of $x \in X$ for which*

$$\sup_n \sum_{k=0}^{n} f \circ T^k(x)$$

is greater than 0.

Then

$$\int_B f d\mu \geq 0.$$

Proof We consider the proof just for the special case of T being invertible.

Let B_1 be the set of x for which $f(x) > 0$, and then at $n > 1$ let B_{n+1} be the set of x for which

$$\sum_{k=0}^{m} f(T^k(x)) \leq 0$$

all $m < n$ but

$$\sum_{k=0}^{n} f(T^k(x)) > 0.$$

It suffices to show that at any n we have

$$\int_{B_1 \cup B_2 \cup \ldots \cup B_n} f d\mu \geq 0.$$

Note that $T[B_\ell] \subset \bigcup_{m<\ell} B_m$ and then $T^k[B_\ell] \subset \bigcup_{m \leq \ell - k} B_m$.

Claim: For $m \leq n$, $i < j < m$ we have

$$T^i[B_m] \cap T^j[B_m] = \emptyset.$$

Proof of Claim: Otherwise choose $x, y \in B_m$ with $T^i(x) = T^j(y)$. Then since T is one to one, we obtain $T^{j-i}(y) = x$. But $x \in B_m$ and by the remarks above we have $T^{j-i} \in B_\ell$ for some $\ell < m$, contradicting disjointness of the sets $B_1, B_2, \ldots..$ (Claim□)

We let $B'_n = B_n$ and recursively define

$$B'_m = B_m \setminus \bigcup_{\ell > m} \bigcup_{k \leq \ell - m} T^k[B'_\ell].$$

Thus

$$\{T^k[B'_\ell] : \ell \leq n, k < \ell\}$$

gives a disjoint covering of $B_1 \cup \ldots \cup B_n$. Moreover

$$\int_{B'_m \cup T[B'_m] \cup \ldots \cup T^{m-1}[B'_m]} f d\mu$$

$$= \int_{B'_m} \sum_{k=0}^{m-1} f \circ T^k d\mu > 0,$$

as required. □

The proof given above does *not* work in the case T non-invertible. For one thing, it is easy to come up with examples where, for instance, $T[B_3]$ and $T^2[B_3]$ are not disjoint. For another, there might be measurable sets where $\mu(T[B]) \neq \mu(B)$, and hence the very last equation of the proof above might not hold true.

I will sketch the ideas behind a proof which does not assume T is invertible.

We first let A_n be the set of x where there is some $m \leq n$ for which $\sum_{k=0}^{m-1} f \circ T^k d\mu > 0$. At each n we let f_n be $f \cdot \chi_{A_n}$ (so that f_n equals f on A_n and zero outside A_n). Again, it suffices to show each $\int f_n d\mu$ greater than zero. By T being measure preserving, we have at each N

$$\int f_n d\mu = \frac{1}{N} \int \sum_{k=0}^{N-1} f_n \circ T^k d\mu.$$

The next tricky point is that for some arbitrary $x \in X$ we can look at the sequence of values

$$f_n(x), f_n(T(x)), f_n(T^2(x)), ...f_n(T^{N-1}(x)).$$

Thinking of $N >> n$ and imagining that most of the time $T^i(x) \in A_n$, we can find $i_i < i_2 < i_3 < ... < i_p < N$ such that between each successive term there are at most n many non-zero values and between any of the successive markers we have

$$\sum_{k=i_\ell}^{i_{\ell+1}-1} f_n(T^k(x)) \geq 0.$$

Then it is not hard to calculate that

$$\begin{aligned} \int f_n d\mu &= \frac{1}{N} \int \sum_{k=0}^{N-1} f_n \circ T^k d\mu \\ &\geq \frac{-n}{N} \int \|f\|_1 d\mu, \end{aligned}$$

which tends to zero as N heads towards infinity.

Corollary 16.11 *Let (X, μ) and T as above. Let $f \in L^1(X, \mu)$ be real valued. Then*

$$\int_{\{x:\sup_n \frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x))>\alpha\}} f d\mu \geq \alpha\mu(\{x : \sup_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) > \alpha\}),$$

$$\int_{\{x:\inf_n \frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x))<\alpha\}} f d\mu \leq \alpha\mu(\{x : \inf_n \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) < \alpha\}),$$

for $\alpha \in \mathbb{R}$.

Proof We only need to convince ourselves of the first equation, since the first implies the second after we replace f by $-f$. But here if we let

$$g = f - \alpha$$

then the first equation amounts to

$$\int_{\{x:\sup_n \frac{1}{n}\sum_{k=0}^{n-1} g(T^k(x))>0\}} g d\mu \geq 0\mu(\{x : \sup_n \frac{1}{n} \sum_{k=0}^{n-1} g(T^k(x)) > 0\}) = 0,$$

and follows from the last lemma. □

Theorem 16.12 *(The pointwise ergodic theorem) Let (X, μ) be a standard Borel probability space. Let $f \in L^1(X, \mu)$ and let T be an ergodic m.p.t. Then at almost every x*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) \to \int f d\mu.$$

Proof It suffices to prove this for f real valued and positive. After linearity extends the result to complex valued functions.

First let us show that almost everywhere this limit exists. For a contradiction suppose there is $\alpha < \beta$ with

$$\text{limsup}\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) > \beta,$$

$$\text{liminf}\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)) < \alpha,$$

on a non-null set of x. By ergodicity we obtain this to be true at almost every x. This in particular gives that at every x

$$\text{sup}_n \sum_{k=0}^{n-1} f(T^k(x)) > \beta$$

and

$$\text{inf}_n \sum_{k=0}^{n-1} f(T^k(x)) < \alpha.$$

Applying 16.11 we obtain $\int f d\mu \geq \beta$ and then $\int f d\mu \leq \alpha$, with a contradiction.

So we obtain some fixed $\alpha \in \mathbb{R}$ with

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) \to \alpha$$

for almost all $x \in X$.

Claim: For each $\epsilon > 0$

$$\int f d\mu - \alpha < \epsilon.$$

Proof of Claim: Let $g : X \to \mathbb{R}$ be a positive, measurable, bounded function with

$$\int |f - g| d\mu < \epsilon$$

and

$$g \leq f.$$

Let $\beta \in \mathbb{R}^{>0}$ be an upper bound on g.

(i) By the argument from the first part of this theorem, we can find some $\gamma \in \mathbb{R}$ such that for almost all $x \in X$

$$\frac{1}{n}\sum_{k=0}^{n-1} g \circ T^k(x) \to \gamma.$$

(ii) Now appealing to the boundedness of g, we have have at each n

$$\frac{1}{n}\sum_{k=0}^{n-1} g \circ T^k < \beta.$$

Thus by dominated convergence, as at 5.4 above, we have

$$\int \frac{1}{n}\sum_{k=0}^{n-1} g(T^k(x))d\mu(x) \to \int \gamma d\mu = \gamma.$$

Since T is measure preserving, we have at each n

$$\int \frac{1}{n}\sum_{k=0}^{n-1} g(T^k(x))d\mu(x) = \int g(x)d\mu(x),$$

and hence

$$\gamma = \int g d\mu.$$

(Aside: Since g is bounded and μ is finite, we have $g \in L^2(X,\mu)$, so we could have also established this last point by appealing to 16.7.)

(iii) Since $g \leq f$ we have at each n

$$\int \frac{1}{n}\sum_{k=0}^{n-1} g \circ T^k d\mu \leq \int \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k d\mu,$$

and hence

$$\gamma = \int \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} g \circ T^k d\mu \leq \int \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k d\mu = \alpha.$$

(iv) Since

$$\int f d\mu - \int g d\mu < \epsilon,$$

or equivalently,

$$\int g d\mu + \epsilon > \int f d\mu,$$

we obtain from (ii) that

$$\gamma + \epsilon > \int f d\mu,$$

and then from (iii) that

$$\alpha + \epsilon > \int f d\mu,$$

as required. (Claim$\square$)

Quantifying over all $\epsilon > 0$ we obtain

$$\alpha \geq \int f.$$

On the other hand, since at each n

$$0 \leq \int \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k = \int f d\mu$$

we have

$$\alpha = \int \mathrm{lim}_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k d\mu \leq \mathrm{limsup}_{n\to\infty} \int \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k d\mu = \int f d\mu,$$

thereby completing the proof of the theorem. □

16.3 Mixing properties

Definition Let T be an m.p.t. on a standard Borel probability space (X, μ). T is said to be *mixing* if for all measurable A, B

$$\mu(T^{-n}[A] \cap B) \to \mu(A)\mu(B)$$

as $n \to \infty$.

So the example we had of this before was the Bernoulli shift. For S a finite set and $X = S^{\mathbb{N}}$ with the product of the counting measure, we let $T(f)$ be the resulting of shifting one place further along:

$$T(f)(n) = f(n + 1).$$

We observed back in in the start of §16.1 that the finite boolean combinations of cylinder sets are dense in the measure algebra and for A and B arising in this class we have $\mu(T^{-n}[A] \cap B)$ actually equal to $\mu(A)\mu(B)$ for all sufficiently large n.

Mixing is sometimes also called *strong mixing* to distinguish it from *weak mixing*, below.

Definition Let T be an invertible m.p.t. on a standard Borel probability space (X, μ). Let $\mathcal{H}$ be the Hilbert space of square integrable functions from X to $\mathbb{C}$. We then define the corresponding unitary operator

$$U_T : \mathcal{H} \to \mathcal{H}$$

by

$$U_T(f)(x) = f(T(x)).$$

Thus f is an eigenvector for U_T if there is some $\alpha \in C$ with

$$f(T(x)) = \alpha f(x)$$

for almost every x.

We say that T is *discrete spectrum* if $\mathcal{H}$ is spanned by the eigenvectors for U_T.

Again we had an example in §16.1. We let T act on $\mathbb{R}/\mathbb{Z}$ by some kind of irrational rotation: For instance

$$T(x) = x + \sqrt{2} \,\mathrm{mod} 1.$$

There are many equivalent definitions of *weak mixing*. I will take one.

Definition Let T be an m.p.t. on a standard Borel probability space (X, μ). T is said to be *weak mixing* if the induced transformation $T \times T : X \times X \to X \times X$

$$(x, y) \mapsto (T(x), T(y))$$

is ergodic with respect to the product measure $\mu \times \mu$.

Weak mixing also has a Hilbert space characterization: T is weak mixing if and only if the corresponding operator U_T has no eigenvectors other than the constant functions. (This is *not* trivial to prove. See [8] for a proof.)

16.4 The ergodic decomposition theorem

It turns out that some kind of rationale can be provided for studying only ergodic transformations: Every transformation can be written as a direct integral of ergodic transformations.

For simplicity let us assume T is an invertible m.p.t. on a standard Borel probability space (X, μ). Consider the measure algebra (M, d) consisting of measurable subsets of X with

$$d(A, B) = \mu(A \Delta B).$$

It is a standard fact that this is a separable metric space (after we take the customary step of identifying sets which agree a.e.).

Let M_0 be those elements of M which are invariant under T. Let $\{A_n : n \in \mathbb{N}\}$ be a countable dense subset of M_0. We then define

$$\pi : X \to 2^{\mathbb{N}}$$

by

$$\pi(x)(n) = 1$$

if $x \in A_n$, and $= 0$ if $x \notin A_n$.

At each $y \in 2^{\mathbb{N}}$ we let $X_y = \pi^{-1}[\{y\}]$, the set of x with $\pi(x) = y$. It follows immediately from the definitions that each X_y is T-invariant. Following the measure disintegration theorem we can find measure ν on $2^{\mathbb{N}}$ and at each y a μ_y concentrating X_y with

$$\mu = \int \mu_y d\nu(y).$$

It is easily verified that T must act in a measure preserving manner on almost every (X_y, μ_y), or we could stitch together the counterexamples, choosing say $A_y \subset X_y$ with $\mu_y(T^{-1}[A_y]) > \mu_y(A_y)$ on a non-null set of y, and take $A = \{x : x \in X_{\pi(x)}\}$ to get a measurable set with $T^{-1}[A]$ having a greater measure than A.[6]

A similar argument serves to show ergodicity. If for some non-null set of y we have T not ergodic on (X_y, μ_y), then we could fix an $k \in \mathbb{N}$ and a non-null set C of y for which there is some $B_y \subset X_y$ with measure between $(\frac{1}{k}, 1 - \frac{1}{k})$. We let B be the set of $\{x : x \in B_{\pi(x)}\}$ and then it is easily checked that for any n

$$d(A_n, B) = \mu(A_n \Delta B) > \frac{1}{k}\mu(C),$$

contradicting density of $\{A_n : n \in \mathbb{N}\}$ in the subalgebra of invariant measurable sets.

[6] There are in fact some subtle points here. We need to know that $y \mapsto A_y$ is suitably measurable for the corresponding definition of A to provide a measurable set. I will ignore these details.

17 Amenability

17.1 Hyperfiniteness and amenability

The main context of the last section was the following: We have a measure preserving transformation

$$T : (X, \mu) \to (X, \mu)$$

on a standard Borel probability space (X, μ). Initially we made no assumption about invertibility, but after a while it was convenient to assume that T is one to one and onto, and then T^{-1} will itself be an m.p.t. Note then that if we let $M(X, \mu)$ be the space of measurable subsets of X with the metric

$$d(A, B) = \mu(A \Delta B)$$

and the identification of sets with $A\Delta B$ null, then T can be viewed as acting by automorphisms on this space. In fact we have an action of the group $\mathbb{Z}$, either on the space X or this derived space of measurable subsets:

$$\ell \cdot x = T^\ell(x);$$
$$\ell \cdot A = T^\ell[A].$$

So this is something like a representation of the group $\mathbb{Z}$. The slightly subtle point is that it is preserving structure not at the level of X, which in its own right comes with no topological or algebraic structure, but at the level of the measurable subsets.

In this section I want to look at the ergodic theory of general groups. Our new context will be this: Γ is some countable group. (X, μ) is a standard Borel probability space equipped with an action by Γ. For each $\gamma \in \Gamma$ the resulting function

$$\gamma \cdot (\cdot) : X \to X,$$
$$x \mapsto \gamma \cdot x$$

will be an m.p.t. There are many different directions we could lead to. The notes below should be viewed as a short introduction to one of the topics in this area.

I am going to structure the discussion around the idea of *hyperfiniteness*. One comparison between action of $\mathbb{Z}$ and more complicated groups such as $\mathbb{F}_2$, the free group on two generators, is that the former give rise to *hyperfinite* orbit equivalence relations. We will prove that free, measure preserving actions of $\mathbb{F}_2$ on standard Borel probability spaces are never hyperfinite, and this in turn can be used to give an application to the theory of *percolation*, in the sense understood by probabilists.

Definition An equivalence relation E on a standard Borel probability space is *Borel* if it is Borel as a subset of $X \times X$. It is *finite* if every equivalence class is finite. It is *countable* if every equivalence class is countable. It is *hyperfinite* if it can be represented as an increasing union of finite Borel equivalence relations – that is to say there are equivalence relations $(E_i)_{i\in\mathbb{N}}$ such that:

(i) $E_i \subset E_{i+1}$;
(ii) each E_i is Borel and finite;
(iii) $E = \bigcup_{i\in\mathbb{N}} E_i$.

Exercise Let Γ be a countable group acting on a standard Borel space by Borel automorphisms – which is to say, for each $\gamma \in \Gamma$

$$x \mapsto g \cdot x$$

is a Borel function. Show that the induced orbit equivalence relation

$$E_\Gamma = \{(x, \gamma \cdot x) : x \in X, \gamma \in \Gamma\}$$

is Borel.

Example Recall the example of an irrational rotation. $X = \mathbb{R}/\mathbb{Z}$. We let T act by

$$T(x) = \sqrt{2} + x \bmod 1.$$

T is a homeomorphism of X to itself, and we obtain an action of $\mathbb{Z}$ with

$$\ell \cdot x = T^{\ell}(x).$$

First of all, the equivalence relation is countable, since every equivalence class,

$$[x] = \{T^{\ell}(x) : \ell \in \mathbb{Z}\},$$

is obviously countable. As for showing it Borel, note that at each ℓ the set

$$R_{\ell} = \{(x, T^{\ell}(x)) : x \in \mathbb{R}/\mathbb{Z}\}$$

is closed, and the induced equivalence relation is then

$$E = \bigcup_{\ell \in \mathbb{Z}} R_{\ell},$$

and hence F_{σ}.

Finally, the equivalence relation is in fact also hyperfinite. At each n let U_n be the open interval $(0, \frac{1}{n})$. I have chosen these so they are decreasing and

$$\bigcap_{n \in \mathbb{N}} U_n = \emptyset.$$

We then set $xE_nT^{\ell}(x)$ if either
(i) $\ell = 0$; or
(ii) $\ell > 0$ and $x, T(x), T^2(x), ...T^{\ell}(x)$ are all *outside* U_n; or
(ii) $\ell < 0$ and $x, T^{-1}(x), T^{-2}(x), ...T^{\ell}(x)$ are all *outside* U_n.

In fact something much more general is true:

Lemma 17.1 *Let $\mathbb{Z}$ act by Borel automorphisms on a standard Borel space X. Then the resulting orbit equivalence relation is hyperfinite.*

The proof is not deep, but there are some minor technical issues I do not want to get caught up making completely precise. The first step is to apply 8.6 simultaneously to all the Borel sets of the form $T^{\ell_1}[V_1] \cap T^{\ell_2}[V_2] \cap ... \cap T^{\ell_n}[V_n]$ where $V_1, ..., V_n$ are open. This will give a stronger Polish topology in which T acts by homeomorphisms.

Now if we are really, really lucky, the resulting orbits, $[x] = \{T^{\ell}(x) : \ell \in \mathbb{Z}\}$, will all be dense and so will each of the *forward orbits*,

$$\{T^{\ell}(x) : \ell \geq 0\},$$

and the *backwards orbits*,

$$\{T^{\ell}(x) : \ell \leq 0\}.$$

Then the proof is much the same as in the example above – I simply choose the open sets with the properties indicated there.

A more complicated case is when the orbits are dense, but not necessarily in both directions. Then we choose for each x some basic open W_x so that $[x]$ has a last or first moment when it meets that open set.

With some care we can do this so that $xEy \Rightarrow X_x = W_y$ and if we let $s(x)$ be that special last or first point, then

$$x \mapsto s(x)$$

is not only $\mathbb{Z}$-invariant but Borel. We then let xE_ky if $x = y$ or for some $i, j \in \{-k, -k+1, ..., 0, 1, ...k\}$ we have $T^j(x) = s(x), T^i(y) = s(x) = s(y)$.

In general there is no guarantee, of course, that the orbits will be dense. The argument for in this more typical case involves decomposing X into Borel subsets on which all points have the same closure and working on each of these components separately. Suffice to say there are technicalities, but they idea is not deep.

It is not presently understood which countable groups give rise to hyperfinite equivalence relations[7]. One of the deepest theorems in this entire area was proved in 2005:

Theorem 17.2 *(Gao-Jackson) Let Γ be a countable abelian group acting by Borel automorphisms on a standard Borel probability space X. Then the resulting orbit equivalence relation E_Γ is hyperfinite.*

Their long proof is still yet to be published.

Definition $\mathbb{F}_2 = \langle a, b \rangle$ is the free group on generators a and b. An element of $\mathbb{F}_2$ consists of *reduced* words in the letters $\{a, a^{-1}, b, b^{-1}\}$ – where *reduced* means there should be no adjacent appearances of a and a^{-1} or b and b^{-1}. We multiply elements of $\mathbb{F}_2$ by concatenating and then reducing.

For instance let $\sigma = ab^2a^{-1}$ and $\tau = ab^3aba^{-1}$. (The usual notational shortcut: I write ab^2a^{-1} instead of $abba^{-1}$.) We multiply by

$$\sigma\tau = ab^5aba^{-1}.$$

Technically speaking, the identity of $\mathbb{F}_2$ is the empty string. We usually denote this by e – as against lyrically just leaving an empty space and hoping it is recognized for its role as the identity of the group.

Definition For Γ a countable group, we let $\ell_1(\Gamma)$ be the space of functions

$$f : \Gamma \to \mathbb{R}$$

with

$$\sum_{\sigma\in\Gamma} |f(\sigma)| < \infty.$$

For $f \in \ell_1(\Gamma)$ we let

$$\|f\| = \sum_{\sigma\in\Gamma} |f(\sigma)|.$$

We let Γ act on $\ell_1(\Gamma)$ by

$$\sigma \cdot f(\tau) = f(\sigma^{-1}\tau).$$

We then say that the group Γ is *amenable* if for any finite $F \subset \Gamma$ and $\epsilon > 0$ there is some $f \in \ell^1(\Gamma)$ with

$$\|f\| = 1$$

and

$$\|f - \sigma \cdot f\| < \epsilon$$

all $\sigma \in F$.

[7]BUT be warned. There is a weaker use of the term *hyperfinite* under which the answer is understood. Some authors take *hyperfinite* to mean, in effect, hyperfinite on a conull set. In this weaker sense, Connes, Feldman, and Weiss, showed that every countable amenable group gives rise to a hyperfinite equivalence relation when it acts measurably on a standard Borel probability space

At the present this will probably seem a rather technical definition. It turns out there are many equivalent formulations of amenability – the definition I have chosen above is neither the best nor the most common, but simply the most convenient for the proofs that lie ahead. For instance amenability is equivalent to the existence of a *finitely additive* Γ-invariant function

$$m : P(\Gamma) \to [0,1]$$

with $m(\Gamma) = 1$. Amenability is also equivalent to the following remarkable property: Whenever Γ acts continuously on a compact metric space, there is a Γ-invariant probability measure. It would take us far to far afield to enter in to the proof that this property characterizes amenability, but it would probably be helpful to see that this property holds in the case of $\mathbb{Z}$.

Theorem 17.3 *Let K be a compact metric space and suppose $\mathbb{Z}$ acts continuously on K. Then there is a $\mathbb{Z}$-invariant probability measure on K.*

Proof Recall from §11 that we use $P(K)$ to denote the Borel probability measures on K in the induced weak star topology, and that in this topology it forms a compact metrizable space.

Begin with any $\nu \in P(K)$. Let

$$\mu_N = \frac{1}{N} \sum_{0 \leq \ell \leq N-1} \ell \cdot \nu,$$

where $\ell \cdot \nu$ is defined by

$$\int f d\ell \cdot \nu = \int f(\ell \cdot x) d\nu(x).$$

Now appealing to the compactness of $P(K)$ we can find a measure μ which is a convergence point of a subsequence of $(\mu_n)_{n \in \mathbb{N}}$. For any $k \in \mathbb{Z}$ and $f \in C(K < \mathbb{R})$ we have

$$\int f d(\mu - k \cdot \mu_N) < \frac{2k}{N} \|f\|,$$

and thus for all $\epsilon > 0$, all $k \in \mathbb{Z}$ there exists an N such that for all $m > N$ and f with $\|f\| \leq 1$ we have

$$|\int f d\mu_m - \int f d(k \cdot \mu_m)| \leq \epsilon.$$

This is a weak star closed condition in the display, and hence we must have for μ that for all $\epsilon > 0$, all $k \in \mathbb{Z}$, and f all with $\|f\| \leq 1$ we have

$$|\int f d\mu - \int f d(k \cdot \mu)| \leq \epsilon.$$

This actually gives that for all $f \in C(K, \mathbb{R})$ and all $k \in \mathbb{Z}$

$$\int f d\mu = \int f d(k \cdot \mu),$$

which amounts to $\mathbb{Z}$-invariance of μ. □

One can also characterize amenability in terms of the existence of left invariant element in the dual of $\ell^\infty(G)$ or in terms of almost invariant unit vectors in the regular representation[8] of G on $\ell^2(G)$.

On it goes. Like I said, the choice I have taken here for defining amenability is technical but convenient to our goals. A discussion of these other characterizations, many of which require subtle ideas from Banach space theory to be proved, can be found in [5],

[8] I.e. induced by the shift in the same way we induced an action of G on $\ell^1(G)$ above

Lemma 17.4 $\mathbb{F}_2$ *is not amenable.*

Proof For $u \in \{a, a^{-1}, b, b^{-1}\}$ we let A_u be the reduced words beginning with u. For $g \in \ell_1(\Gamma)$ we define g_u by

$$g_u(\sigma) = g(\sigma)$$

if $\sigma \in A_u$, and $g_u(\sigma) = 0$ otherwise.

Thus f is the sum of $f_a, f_{a^{-1}}, f_b, f_{b^{-1}}$, as well as its value on e. Note moreover that if $w \neq u^{-1}$, $w, u \in \{a, a^{-1}, b, b^{-1}\}$, then

$$w \cdot A_u \subset A_w.$$

We will take as our F the set $\{a, a^{-1}, b, b^{-1}\}$ and as our ϵ the value $\frac{1}{4}$. Let $f \in \ell_1(\Gamma)$ with $\|f\| = 1$. We will show there is some $\sigma \in F$ with

$$\|f - \sigma \cdot f\| \geq \frac{1}{4}.$$

First choose $u \in F$ with

$$\|f_u + f_{u^{-1}}\| \leq \frac{1}{2}.$$

For $\sigma \notin A_{u^{-1}}$ we have $u \cdot \sigma \in A_u$ and $(u \cdot f)(u \cdot \sigma) = f(\sigma)$. This yields

$$\|(u \cdot f)_u\| \geq \|f - f_{u^{-1}}\|.$$

Similarly

$$\|(u^{-1} \cdot f)_{u^{-1}}\| \geq \|f - f_u\|.$$

One of $\|(u \cdot f)_u\|$ and $\|(u^{-1} \cdot f)_{u^{-1}}\|$ is therefore at least $\frac{3}{4}$ given

$$\|f_u + f_{u^{-1}}\| = \|f_u\| + \|f_{u^{-1}}\| \leq \frac{1}{2}.$$

This gives either

$$\|u \cdot f - f\| \geq \frac{1}{4}$$

or

$$\|u^{-1} \cdot f - f\| \geq \frac{1}{4}.$$

□

Lemma 17.5 $\mathbb{Z}$ *is amenable.*

Proof Let

$$A_n = \{-n, -n+1, ..., 0, 1, ..., n-1, n\}.$$

For $\ell \in Z$ and $A \subset \mathbb{Z}$ we let $\ell \cdot A = \{\ell + k : k \in A\}$. It is then easily seen that

$$\frac{|A_n \Delta \ell \cdot A_n|}{|A_n|} \to 0$$

as $n \to \infty$. Thus if we let

$$f_n = \frac{1}{|A_n|} \chi_{A_n}$$

then each $\|f_n\| = 1$ and

$$\|f_n \ell \cdot f_n\| \to 0$$

as $n \to \infty$. Thus given any finite $F \subset \mathbb{Z}$ and $\epsilon > 0$ we have

$$\|f_n \ell \cdot f_n\| < \epsilon$$

all $\ell \in F$ and n sufficiently large. □

Theorem 17.6 *Let $\mathbb{F}_2$ act freely and by Borel automorphisms on a standard Borel probability space (X, μ). Then the resulting orbit equivalence relation*

$$E_{\mathbb{F}_2} = \{(x, \sigma \cdot x) : \sigma \in \mathbb{F}_2\}$$

is not hyperfinite.

Proof Suppose instead for a contradiction

$$E_{\mathbb{F}_2} = \bigcup_{i \in \mathbb{N}} E_i$$

where the E_i's are finite Borel equivalence relations with $E_i \subset E_{i+1}$. Then at each $i \in \mathbb{N}, x \in X$ we let

$$f_{i,x}(\sigma) = \frac{1}{|[x]_{E_i}|}$$

if $x E_i \sigma^{-1} \cdot x$, and $= 0$ otherwise. (Here $|[x]_{E_i}|$ denotes the number of points E_i-equivalent to x.) Let

$$f_i(\sigma) = \int f_{i,x}(\sigma) d\mu.$$

Since the action is free we obtain each

$$f_{i,x} \in \ell_1(\mathbb{F}_2)$$

with $\|f_{i,x}\| = 1$. Interchanging integration with summation yields

$$\|f\| = \sum_{\sigma \in \mathbb{F}_2} \int f_{i,x}(\sigma) d\mu$$

$$= \int \sum_{\sigma \in \mathbb{F}_2} f_{i,x}(\sigma) d\mu$$

$$= \int \|f_{i,x}\| d\mu = 1.$$

Claim: For each $\sigma \in \mathbb{F}_2$

$$\lim_{i \to \infty} \|f_i - \sigma \cdot f_i\| \to 0.$$

Proof of Claim: Let $A_i = \{x : x E_i \sigma \cdot x\}$. Since $\bigcup_{i \in \mathbb{N}} E_i = E_{\mathbb{F}_2}$ we obtain

$$\mu(A_i) \to \infty$$

as $i \to \infty$. Moreover if $x \in A_i$ then

$$f_{i,\sigma \cdot x}(\sigma^{-1} \tau) = f_{i,x}(\tau)$$

for any $\tau \in \mathbb{F}_2$. Thus

$$\int_{\sigma \cdot A_i} \sigma \cdot f_{i,x}(\tau) d\mu = \int_{\sigma \cdot A_i} f_{i,x}(\sigma^{-1}\tau) d\mu$$

$$= \int_{A_i} f_{i,x}(\tau) d\mu,$$

since σ acts in a measure preserving manner. Thus

$$\|f - \sigma \cdot f\| = \sum_{\tau \in \mathbb{F}_2} (|\int f_{i,x}(\tau) d\mu - \int f_{i,x}(\sigma^{-1}\tau) d\mu|)$$

$$\leq \sum_{\tau \in \mathbb{F}_2} (|\int_{A_i} f_{i,x}(\tau) d\mu - \int_{\sigma \cdot A_i} f_{i,x}(\sigma^{-1}\tau) d\mu|) + \sum_{\tau \in \mathbb{F}_2} (\int_{X \setminus A_i} f_{i,x}(\tau) d\mu + \int_{X \setminus \sigma \cdot A_i} f_{i,x}(\sigma^{-1}\tau) d\mu$$

$$= (\int_{X \setminus A_i} \sum_{\tau \in \mathbb{F}_2} f_{i,x}(\tau) d\mu + \int_{X \setminus \sigma \cdot A_i} \sum_{\tau \in \mathbb{F}_2} f_{i,x}(\sigma^{-1}\tau) d\mu$$

$$= \mu(X \setminus A_i) + \mu(X \setminus \sigma \cdot A_i.$$

Since σ acts in a measure preserving manner, this in turn equals

$$2\mu(X \setminus A_i),$$

which goes to 0 as $i \to \infty$. (Claim□)

But now for any finite $F \subset \mathbb{F}_2$ and $\epsilon > 0$ we will have at all sufficiently large i

$$\forall \sigma \in F(\|f_i - \sigma \cdot f_i\| < \epsilon),$$

with a contradiction to non-amenability of $\mathbb{F}_2$. □

All we used about $\mathbb{F}_2$ is its non-amenability. Thus the proof shows:

Theorem 17.7 *Let Γ be a non-amenable group acting freely and by measure preserving transformations on a standard Borel probability space (X, μ). Then the resulting orbit equivalence relation is not hyperfinite.*

As a corollary to this theorem we obtain another proof that $\mathbb{Z}$ is amenable.

17.2 An application to percolation

Let $G = (V, E)$ be an infinite connected graph. Imagine we have some random process which will reduce the graph, leaving some edges in while erasing many others. Let p be a real number between 0 and 1. At each edge $c \in E$ we suppose our random process independently gives the edge p chance of remaining in the graph. Roll the dice and conduct the experiment, and at the end, after all these edges have had their chance, we will be left with a subgraph of G, and we can ask various kinds of qualitative questions: Whether there is an infinite connected component[9], and if so, how many.

This is called a *percolation* problem. Certain kinds of probabilists and physicists are interested in which kinds of graphs will have a $p < 1$ for which the resulting experiment is certain to leave us with at least one infinite component. We can use the ideas from the last subsection to show that with the *Cayley graph* of $\mathbb{F}_2$ there is a $p < 1$ for which the above experiment is almost certain to lead to an infinite component.

Here is a purely mathematical way to formulate the problem.

Definition Let $G = (V, E)$ be a countable graph. Let $X(G)$ be the space of all functions from E to (0, 1),

$$E^{(0,1)} = \prod_E (0, 1),$$

[9]Recall: If $H = (W, F)$ is a graph and $w \in W$, the *connected component of* w is the set of all $v \in W$ for which there is a path $w_0 = w, w_1, ..., w_n = v$ with each $\{w_i, w_{i+1}\} \in F$

equipped with the product topology and the infinite product of Lebesgue measure on (0, 1). (Note that $X(G)$ is a standard Borel probability space.) For each $f \in X(G)$ and $p \in [0,1]$ we let $G_{f,p} = (V, E_{f,p})$, where V is as before but

$$E_{f,p} = \{c \in E : f(c) > p\}.$$

So from the *experiment* f and the *probability* p we have obtained a randomly presented subgraph of G.

We then let $p_c(G)$ be the least $q \in [0,1]$ such for a non-null set of $f \in X(G)$ we have an infinite connected component in $G_{f,p}$ whenever $p \in (0,1)$ has $p > q$.

There is something rather devious in the way I have phrased the definition. If there is *no* $p \in (0,1)$ for which there is a non-zero chance of $G_{f,p}$ having an infinite component, then p_c gets set to the default value of 1.

Definition Let Γ be a countable group and S a generating set – which is to say that every element of Γ can eventually be obtained by multiplying together elements of S and their inverses. We then let the induced *Cayley graph*, $G(\Gamma, S)$ be the graph with vertex set Γ and an edge running between σ and τ if for some $s \in S \cup S^{-1}$ we have

$$\sigma s = \tau.$$

We let Γ act on $G(\Gamma, S)$ by

$$\gamma \cdot \sigma = \gamma\sigma,$$

$$\gamma \cdot \{\sigma, \sigma s\} = \{\gamma\sigma, \gamma\sigma s\}.$$

Exercise Let $G = (V, E)$ be the Cayley graph of $\mathbb{Z}$ with the generating set $S = \{1\}$. Show that $p_c(G) = 1$.

Theorem 17.8 *Let $\mathbb{F}_2 = \langle a, b\rangle$ and take as our generating set $S = \{a, b\}$. Then*

$$p_c(G(\mathbb{F}_2, S)) < 1.$$

Proof Write $G(\mathbb{F}_2, S) = G = (V, E)$. Hence

$$V = \mathbb{F}_2$$

and E is the set of all

$$\{\sigma, \sigma u\},$$

where $\sigma \in \mathbb{F}_2, u \in \{a, a^{-1}, b, b^{-1}\}$. At every $p \in (0,1)$ the set

$$A_p = \{f \in X(G) : G_{f,p} \text{ has an infinite component}\}.$$

Claim: Each A_p is Borel.

Proof of Claim: Fix p. Given $\gamma, \tau \in \mathbb{F}_2$ we let

$$B_{\gamma,\tau} = \{f : \gamma, \tau \text{ connected in } G_{f,p}\}.$$

This set is Borel, since there is a unique loopless path

$$\gamma, \gamma u_1, \gamma u_1 u_2, ..., \tau$$

from γ to τ, and then $f \in B_{\gamma,\tau}$ if and only if f assumes a value greater than p at each of the needed edges. Thus if we $(\tau_n)_{n\in\mathbb{N}}$ enumerate the free group, and set

$$C_\gamma = \bigcap_{m\in\mathbb{N}} \bigcup_{n>m} B_{\gamma,\tau_n},$$

then C_γ is seen to be Borel. Note that C_γ is the collection of f's for which γ has an infinite component in $G_{f,p}$.

Finally

$$A_p = \bigcup_{\gamma \in \mathbb{F}_2} C_\gamma.$$

(Claim□)

Let us assume for a contradiction that A_p is null.

We let $\mathbb{F}_2$ act on $X(G)$ by pivoting through its action on the Cayley graph. Given $f \in X(G)$ and $\sigma \in \mathbb{F}_2$ we define $\sigma \cdot f$ by

$$\sigma \cdot f(\{\tau, \tau u\}) = f(\{\sigma^{-1}\tau, \sigma^{-1}\tau u\}).$$

It is easily checked that the action of $\mathbb{F}_2$ on $X(G)$ is measure preserving and each A_p is $\mathbb{F}_2$-invariant – in the sense that $\gamma \cdot A_p = A_p$ for all $\gamma \in \mathbb{F}_2$.

Let

$$Y = \bigcup_{n \in \mathbb{N}} (X(G) \setminus A_{1-\frac{1}{n}}).$$

This is a conull, Borel, $\mathbb{F}_2$-invariant subset of $X(G)$. Hence it is a standard Borel probability space on which $\mathbb{F}_2$ acts by measure preserving transformations. It is then an easy exercise to check that for any $\sigma \in \mathbb{F}_2$, $\sigma \neq e$, the collection of $f \in Y$ for which $\sigma \cdot f \neq f$ is again non-null.

Finally we then come to

$$X = \{f \in Y : \forall \gamma \in \mathbb{F}_2 (\gamma \neq e \Rightarrow \gamma \cdot f \neq f)\},$$

a standard Borel probability space on which $\mathbb{F}_2$ acts freely.

We let E_n be the set of $(f, \gamma \cdot f)$ such that e is connected to γ in $G_{f,1-\frac{1}{n}}$. Digesting the definition, this comes out the same as asking γ^{-1} be connected to e in $G_{\gamma \cdot f,1-\frac{1}{n}}$, and we see that this defines a Borel equivalence relation. The finiteness of the components on the various $G_{f,1-\frac{1}{n}}$ entails that each E_n is finite. For each $\tau \in \mathbb{F}_2, u \in \{a, a^{-1}, b, b^{-1}\}$ and $f \in X$ we have

$$\{\tau, \tau u\} \in A_{1-\frac{1}{n}}$$

all sufficiently large n.

Hence on X

$$E_{\mathbb{F}_2} = \bigcup_{n \in \mathbb{N}} E_n,$$

contradicting 17.6. □

References

[1] A.A. Borovkov, **Probability Theory,** OPA, Amsterdam, 1998.

[2] A. Connes, J. Feldman, B. Weiss, *An amenable equivalence relation is generated by a single automorphism,* **Journal of Ergodic Theory and Dynamical Systems,** vo. 1(1981), pp. 431-450.

[3] J.B. Conway, **A Course in Functional Analysis**, Springer-Verlag, New-York, 1990.

[4] T.W. Gamelin, R.E. Greene, **Introduction to Topology,** Dover, New York, 1999.

[5] F.P. Greenleaf, **Invariant means on topological groups and their applications,** Van Nostrand Mathematical Studies, No. 16 Van Nostrand Reinhold Co., New York-Toronto, London 1969.

[6] A. S. Kechris, **Classical Descriptive Set Theory,** Springer-Verlag, New-York, 1995.

[7] J. J. Koliha, **620-312 Linear Analysis,** course notes. Available publicly as **Metrics, Norms and Integrals: An Introduction to Contemporary Analysis,** World Scientific, Singapore, 2008

[8] K. Petersen, **Ergodic Theory,** Cambridge University Press, Cambridge, 1983.

[9] H.L. Royden, **Principles of Mathematical Analysis**, McGraw-Hill Book Company, New York, 1976.

[10] W. Rudin, **Real and Complex Analysis,** Walter Rudin, McGraw-Hill, New York, 1966.

[11] R.M. Solovay, *A model of set-theory in which every set of reals is Lebesgue measurable,* **Annals of Mathematics,** (2) 92 1970 1–56

[12] S.M. Srivastava, **A Course on Borel Sets**, Springer-Verlag, New York, 1998.