Measure Theory, 2008

May 26, 2008

I will assume familiarity with the basic measure from chapters 12 and 13 of [6], though a brief summary is given below. [8] provides a sweeping general reference for measure theory and analysis. The appendix of [1] covers many of the essential facts we will need, specifically tailored for the point of view of a probabilist. Later in the course we will explore more advanced topics in ergodic theory, and here one could refer to [7]. As we push on I will attempt to update the references.

These notes are a work in progress. Let me know about any errors or typos and I can issue a corrected version.

1 Review and general remarks

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

Exercise The Borel subsets of $\mathbb{R}^N$ 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.

Definition If S is a set and $\Sigma \subset \mathcal{P}(S)$ a σ-algebra, then

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

is a *measure* if

(a) $\mu(\emptyset) = 0$;
(b) $\mu(A) \geq 0$ all $A \in \Sigma$;
(c) $\mu(\bigcup_i A_i)$ equals

$$\sum_{i=1}^{\infty} \mu(A_i)$$

whenever $(A_i)_{i \in \mathbb{N}}$ is a sequence of disjoint elements of Σ.

We will take without proof the following key fact from [6].

Theorem 1.1 *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 unique measure*

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

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

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

Definition The unique measure described by the above theorem is called *the Lebesgue measure.*

It would take us far too afield to prove the theorem here, and instead let us just agree that it has been dealt with in [6]. However it is worth mentioning a key underlying fact which can be used to give its proof.

Lemma 1.2 *If B is a Borel subset of $\mathbb{R}^N$ and $\epsilon > 0$ then there are rectangles $(A_i)_{i\in\mathbb{N}}, (B_i)_{i\in\mathbb{N}}$ such that*
(a) $B \subset \bigcup A_i$;
(b) $\bigcup A_i \setminus B \subset \bigcup B_i$;
(c) $\sum_{i\in\mathbb{N}} m(B_i) < \epsilon$.

If you don't already know the proof, I will leave it as an exercise, however there is something important about the method involved which you should alerted to. We let Σ^* be the collection of all sets with properties (a)-(c). The method is then to show that Σ^* includes the open sets (obvious), is closed under complements (needs a bit of thought), and closed under countable unions (easy). Thus Σ^* is a σ-algebra containing the open sets – since the Borel sets form the *smallest* σ-algebra containing the open sets, they must be included in Σ^*.

Definition We say that $M \subset \mathbb{R}^N$ is *measurable* if there are Borel sets A, B with

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

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

We then let $m(M) = m(A)$. This extension of m to this class of measurable sets is said to be the *completion* of m. It is easily checked that the measurable sets form a σ-algebra and that this completion of m, which I am simply again denoting by m, is again a measure.

Well and good.

However there are other kinds of spaces and objects to which we might wish to assign something like a measure. Before embarking on the formal development, here are some examples.

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:

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

(iv) An example which will be important later on is the *infinite run*, $\{H, T\}^{\mathbb{N}}$. Note that this is a compact metric space in the product topology. For a *cylinder set* of the form $A = \{f : f(1) = S_1, f(2) = S_2, ...f(N) = S_N\}$ we let $\mu(A) = 2^{-N}$; more generally given $S \subset 2^N$ we let $\mu(\{f : f|_N \in S\}) = \frac{|S|}{2^{-N}}$. Since the function μ is σ-additive on the algebra generated by the cylinder sets, it follows from 15.3.8 of [6] that there is a unique extension of μ to a measure on the σ-algebra it generates – which is in fact the full collection of Borel subsets of $\{H, T\}^{\mathbb{N}}$.

(iv) 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 tickling a dangerous African land animal with a long feather or 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 (e.g. the decibels of the dangerous animal's laugh or 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.

Definition A set X equipped with a σ-algebra $\Sigma \subset \mathcal{P}(X)$ is said to be a *measure space.* If $\mu : \Sigma \to \mathbb{R}$ is a measure, then we say that $M \subset X$ is *measurable* (with respect to μ) if there are $A, B \in \Sigma$ with

$$A \subset M \subset B$$

with $B \setminus A$ null (i.e. $\mu(B \setminus A) = 0$). A function $f : X \to \mathbb{R}$ is said to be *measurable with respect to* Σ if for any open set $U \subset \mathbb{R}$ we have $f^{-1}[U] \in \Sigma$. A function $f : X \to \mathbb{R}$ is said to be *measurable with respect to* μ if for any open $U \subset \mathbb{R}$ we have $f^{-1}[U]$ measurable with respect to μ.

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

Lemma 1.3 *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 commit 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 1.4 *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 1.5 *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,*

$$\mu : \Sigma^* \to \mathbb{R},$$

is a measure.

Proof 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 Lebesgue measure on $\mathbb{R}$ is σ-finite.

So much for the brief review of measure. Now for integration.

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

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

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 $a_i \neq 0$.

If, as in [6], we adopt the convention that $\infty \cdot 0 = 0$ then we can simply write this out as

$$\int_X h d\mu = \sum_{i \leq n} \mu(A_i) \cdot a_i.$$

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

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

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 $f^- = \frac{1}{2}(|f| - f)$.

This definition of integration extends 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 $f = \mathrm{Re}f + i\mathrm{Im}f$, where $\mathrm{Re}f : X \to \mathbb{R}$ and $\mathrm{Im}f : X \to \mathbb{R}$ are the real and imaginary parts. We say that f is *integrable* if both these functions are integrable in our earlier sense and let

$$\int_X f d\mu = \int_X \mathrm{Re}f d\mu + i \int_X \mathrm{Im}f d\mu.$$

Real integration. Complex integration. But there is no reason to stop here. We can integrate functions into almost any kind of vector space for which there is an appropriate notion of convergence. Thinking about the process above, it is immediately clear how to integrate a step function, as long as we can take linear combinations on the other side; after that we try to approximate a measurable f by taking suitable simple functions.

Definition Let $\mathbb{B}$ be a Banach space and (X, Σ, μ) a measure space equipped with measure $\mu : \Sigma \to \mathbb{R}$.

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

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

$$\int_X f d\mu$$

be the sum in $\mathbb{B}$ of all $\mu(A_i) \cdot a_i$ for $a_i \neq 0$. We define the L^1-norm of an integrable simple function h by

$$\|h\|_{L^1} = \int_S \|h(.)\| d\mu.$$

Note then that $\|h_1 - h_2\| < \epsilon \Rightarrow |\int_S h_1 d\mu - \int_S h_2 d\mu| < \epsilon$. A function $f : S \to X$ is *Bochner integrable* if there is a sequence (h_n) of integrable simple functions such that

$$\|h_n(x) - f(x)\| \to 0 \text{ for almost all } x \in S,$$

and such that (h_n) is Cauchy with respect to the L^1-norm. The integral of f is then defined as

$$\int_S f \, d\mu = \lim_{n\to\infty} \int_S h_n \, d\mu;$$

The theory of measurable spaces and measurable 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

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

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 1.6 *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 [5]. 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

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

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 1.7 *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 [5].

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 1.8 *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 1.9 *$f : X \to \mathbb{R}$ 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. □

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.

2 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. There are a few lemmas which make it possible for us to analyze the difference between two measures; we will prove these, then the Radon-Nikodym theorem, and finish with an application to conditional expectation. It all starts with the introduction of a variation on the idea of measure.

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. As I have previously remarked, we will not be losing much to always think of our measures as defined on standard Borel spaces, though in the next couple of lemmas we will not be making such an assumption; it will be technically convenient to develop the results in full generality.

Lemma 2.1 *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*

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

Proof First some cosmetic rearrangement. Let $C_n = \bigcap_{i\leq n} B_i$. 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

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

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

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

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

Theorem 2.2 *(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 $B \in \Sigma$, $B \subset A$*

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

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

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

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 no 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 $\mu(A) = \lim_{n\to\infty}\mu(\bigcup_{i\geq n} C_i) = \delta$.

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 2.3 *(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 2.2: 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 2.4 *(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 2.2 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 2.5 *Let $\Sigma_0 \subset \Sigma_1$ be two σ-algebras on a set X. Let μ be a σ-finite measure on (X, Σ_1) and 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 2.4 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. □

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.

3 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 3.1 *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 3.2 *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 $B_i^2 = \{y \in A : d(x_i, y) < 1/2\}$ 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 $B_i^\ell = \{y \in A : d(x_i, y) < 2^{-\ell}\}$ 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. □

Theorem 3.3 *(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 3.1 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 3.2. □

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 3.4 *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 $B_\ell \subset \overline{f[C_{s_\ell}]}$ 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 3.5 *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 3.6 *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.*

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 3.7 *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 3.8 *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 3.4 and 3.7. □

The corollary 3.8 or the next result below is sometimes called *Jankov von Neumann.*

There is a little bit more we can extract from the proof of 3.8 and 3.4. 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 3.9 *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 3.7 we may assume f is continuous. Going through the proof of 3.4, 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

$$C_{\langle n_0^y, n_1^y, \ldots, n_\ell^y \rangle} \subset U.$$

□

Danger, danger: 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 3.10 *(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 [5].

4 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.

Theorem 4.1 Monotone Convergence Theorem *Let (f_n) be a sequence of measurable functions on a standard Borel probability space (X, μ). Assume the functions are non-negative and non-decreasing:*

$$\forall n \forall x \quad 0 \leq f_n(x) \leq f_{n+1}(x).$$

Assume the functions converge pointwise to some function f:

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

as $n \to \infty$.

Then f is measurable and

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

Proof It is a routine consequence of the measurable sets forming a σ-algebra that f will be measurable: For I the interval (a, ∞) we have

$$f^{-1}[I] = \bigcup_{n \in \mathbb{N}} f_n^{-1}[I],$$

and then one just verifies that the Borel sets are generated as a σ-algebra by intervals of this form.

Case 1: $\int f_n d\mu = \infty$. Then we can find a sequence of disjoint measurable sets, (E_i), and positive reals a_i, such that

$$\sum \mu(E_i) \cdot a_i = \infty$$

while $f(x) > a_i$ on each E_i. Then for each i and $x \in E_i$ there will be some n_x such that $f_{n_x}(x) > a_i$; and from this it follows that there will be some m_i such that $\{x \in E_i : n_x \leq m_i\}$ has measure at least $\mu(E_i)/2$. Then if we let N_i be the maximum of m_j, $j \leq i$, then

$$\int f_{N_i} d\mu > \frac{1}{2} \sum_{j \leq i} \mu(E_j) \cdot a_j.$$

Case 2: $\int f_n d\mu < \infty$.

Fix $\epsilon > 0$. At each n let A_n be the set of x for which $|f(x) - f_n(x)|$ is less than or equal to $\epsilon/2$. (Exercise: These sets are measurable) These are increasing sets with

$$\bigcup_n A_n = X,$$

and so eventually we can find some N with

$$\int_{X-A_N} f(x)d\mu < \epsilon/2,$$

and then $\int f d\mu - \int f_N d\mu < \epsilon$. □

Theorem 4.2 Dominated Convergence Theorem *Let (f_n) be a sequence of measurable functions on a standard Borel probability space (X, μ) converging pointwise to some function f. Let g be a non-negative integrable function on X with*

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

for all $n \in \mathbb{N}, x \in X$. Then

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

Proof As in the last argument, we let let A_n be the set of x for which $|f(x) - f_n(x)|$ is less than or equal to $\epsilon/2$. Eventually we come to some N with

$$\int_{X \setminus A_N} g(x)d\mu < \epsilon/4$$

$$\therefore |\int_{X \setminus A_N} f - f_N d\mu| \leq 2 \int_{X \setminus A_N} g d\mu < \epsilon/2$$

and then $|\int f d\mu - \int f_N d\mu| \leq |\int_{A_N}(f - f_N)d\mu| + \int_{X \setminus A_N} g d\mu < \epsilon$. □

On to product spaces. There will be a little bit of notation to clean up.

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 4.3 *$X \times Y$ 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 4.4 *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. $\square$

Lemma 4.5 *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: This is be proved along the lines of 1.2.

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 $C \subset D_n^0$, $X \setminus C \subset D_n^1$,

$$\int (x \mapsto \nu((D_n^0)_x \cap (D_n^1)_x)) d\mu < \rightarrow 0$$

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. □

Lemma 4.6 *Let (X, μ) and (Y, ν) be standard Borel probability spaces. If we define m on $X \times Y$ by*

$$m(B) = \int_X (x \mapsto \nu(B_x))\mu$$

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 4.2 or 4.1 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 4.7 *Let (X, μ) and (Y, ν) be standard Borel probability spaces. Then for any Borel $B \subset X \times Y$*

$$\int_X (x \mapsto \nu(B_x))\mu = \int_Y (y \mapsto \mu(B^y))\nu.$$

Proof Let us first define m as in 4.6 and then likewise m^* by

$$m^*(B) = \int_Y (y \mapsto \mu(B_y))\nu.$$

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 [6], though it is not hard to see directly) that $m = m^*$. □

Definition We let $\mu \times \nu$ be the measure m obtained by either

$$m(B) = \int_X (x \mapsto \nu(B_x))\mu$$

or

$$m(B) = \int_Y (y \mapsto \mu(B^y))\nu.$$

Summarising the above:

Proposition 4.8 *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 4.9 *(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 4.8. 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

$$0\le f_i(x)\le f_{i+1}(x)\le f(x)$$

and $f_n(x)\to f(x)$ at every x. Now the result follows from 4.1. □

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

$$\mu(A) = \int_Z (z \mapsto \mu_z(A \cap Y_z)) d\lambda.$$

5 The Riesz representation theorem

This section will closely follow the treatment in [8].

Definition Let (K, d) be a compact metric space. $C(K)$ consists of all continuous functions

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

For $f \in C(K)$ we let $\|f\|$ be

$$\sup_{x\in K}|f(x)|.$$

Given $r \in \mathbb{R}$ and $f \in C(K)$ we define rf by pointwise multiplication,

$$(rf)(x) = r((f(x)),$$

and similarly for $f, g \in C(K)$ 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, any continuous function on a compact metric space is *uniformly continuous*; in other words

$$\forall\epsilon > 0\exists\delta > 0\forall x, y \in K(d(x, y) < \delta \Rightarrow |f(x) - f(y)| < \epsilon).$$

Then given a sequence $(f_n)_n$ in $C(K)$ 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 of undergraduate analysis[1] that the pointwise limit of a uniformly convergent sequence of uniformly continuous functions is again uniformly continuous, and thus $f \in C(K)$.

In conclusion then we have that $C(K)$ 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 $C(K)$ 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, $C(K)^*$, the *dual space*, is the collection of functions

$$\Lambda : C(K) \to \mathbb{R}$$

that are continuous and *linear*: $\Lambda(f + g) = \Lambda(f) + \Lambda(g)$, $\Lambda(rf) = r\Lambda(f)$. We then let $\|\Lambda\|$ be

$$\sup_{\{f\in C(K):\|f\|=1\}}|\Lambda(f)|.$$

(It is a routine exercise to verify $\|\Lambda\|$ finite from Λ continuous.)

[1] But absolutely let me know if this does not sound familiar to you

We say that $f \in C(K)$ 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 $\Lambda \in C(K)^*$ is *positive* if $\Lambda(f) \geq 0$ whenever f is positive.

Definition Let K be a compact metric space and $f \in C(K)$. 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 5.1 *If K is a compact metric space and $C \subset V$ with C closed and V open, then there is a $f \in C(K)$ 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

$$f(x) = \min(g(x), 1).$$

□

Lemma 5.2 *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 5.1 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 5.3 *Let K be a compact metric space and suppose $\Lambda \in C(K)^*$ 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 $f \in C(K)$.

Proof For V open define $\mu(V)$ to to be

$$\sup_{f \prec V} \Lambda(f).$$

For $B \subset K$ Borel we let $\mu(B)$ be

$$\inf_{V \supset B, V \text{ open}} \mu(V).$$

We let Σ be the collection of all Borel B for which

$$\mu(B) = \sup_{C \subset B, C \text{ closed}} \mu(C).$$

Immediately we have every closed set in C. Our main battle will be to show that Σ includes every Borel set. This in turn is requires a sequence of tactical skirmishes.

Claim 1: If V_1, V_2 are open, then

$$\mu(V_1 \cup V_2) \leq \mu(V_1) + \mu(V_2).$$

Proof of claim: Let $g \prec V_1 \cup V_2$. Applying 5.2 to the support of g we can find h_1, h_2 with

$$0 \leq h_1, h_2 \leq 1,$$

$$h_i \prec V_i,$$

$$h_1(x) + h_2(x) = 1$$

all x in the support of g. Letting $g_i = h_i g$ we have

$$g = g_1 + g_2,$$

$$g_i \prec V_i,$$

and then

$$\Lambda(g) = \Lambda(g_1) + \Lambda(g_2) \leq \mu(V_1) + \mu(V_2).$$

(Claim□)

Clearly then we also obtain $\mu(V_1 \cup V_2 \cup ... \cup V_n) \leq \mu(V_1) + \mu(V_2) + ... + \mu(V_n)$ whenever $V_1, V_2, ...V_n$ is a sequence of open sets.

Claim 2: If $(B_i)_i$ is a sequence of Borel sets, then

$$\mu(\bigcup_{i=1}^{\infty} B_i) \leq \sum_{i=1}^{\infty} \mu(B_i).$$

Proof of claim: Fix $\epsilon > 0$. We take open sets with $(V_i)_i$ with

$$B_i \subset V_i,$$

$$\mu(B_i) > \mu(V_i) - 2^{-i}\epsilon.$$

Letting $V = \bigcup_i V_i$, we have $V \supset \bigcup_i B_i$. If $f \prec V$ then by compactness

$$f \prec \bigcup_{i<N} V_i,$$

some N, whence by the last lemma

$$\Lambda(f) \leq \mu(\bigcup_{i<N} V_i) \leq \sum_{i<N} \mu(V_i) \leq \sum_{i=1}^{\infty} \mu(V_i)$$

$$\leq \epsilon + \sum_i \mu(B_i).$$

(Claim□)

Claim 3: If C closed then

$$\mu(C) = \inf_{C \prec f} \Lambda(f).$$

Proof of Claim: Given an open V with $V \supset C$ and $\mu(V) < \mu(C) + \epsilon$, 5.1 gives us some f with $C \prec f \prec V$, and so

$$\Lambda(f) \leq \mu(V) < \mu(C) + \epsilon.$$

Conversely, given $C \prec f$ and $\epsilon > 0$ we may find open $U \supset C$ with

$$f(x) > 1 - \epsilon$$

all $x \in U$. Then for any $f_0 \prec U$ we have $f_0(x) - f(x) < \epsilon$ at every x, and hence

$$\Lambda(f_0) < \Lambda(f) - \epsilon.$$

(Claim□)

Claim 4: Every open set is in Σ.

Proof of Claim: Let V be open and $\epsilon > 0$. Let $f \prec V$ with

$$\Lambda(f) > \mu(V) - \epsilon.$$

Then letting C be the support of f we obtain

$$\mu(C) \geq \Lambda(f) > \mu(V) - \epsilon.$$

(Claim□)

Claim 5: If C_1, C_2 are disjoint and closed, then

$$\mu(C_1 \cup C_2) = \mu(C_1) + \mu(C_2).$$

Proof of Claim: $\leq$ follows from claim 2.

For $\geq$ we apply the criterion of claim 3 and suppose

$$C_1 \cup C_2 \prec g.$$

Applying 5.2 we get f_1, f_2 with

$$f_1 + f_2 = 1,$$
$$f_1|_{C_2} = 0,$$
$$f_2|_{C_1} = 0.$$

$C_i \prec f_i g$ and $g = f_1 g + f_2 g$. (Claim□)

Claim 5 then extends to finite unions of disjoint closed sets.

Claim 6: If $(B_i)_i$ are disjoint sets in Σ and $B = \bigcup_i B_i$, then

$$\mu(B) = \sum_i \mu(B_i)$$

and $B \in \Sigma$.

Proof of Claim: (i) $\mu(B) \geq \sum_i \mu(B_i)$: Let $\epsilon > 0$. Take closed $C_i \subset B_i$ with

$$\mu(C_i) > \mu(B_i) - \epsilon 2^{-i}.$$

Claim 5 gives
$$\sum_{i<N} \mu(C_i) = \mu(\bigcup_{i<N} C_i)$$
for every finite N. Thus
$$(\sum_{i<N} \mu(B_i)) - \epsilon < \sum_{i<N} \mu(C_i) = \mu(\bigcup_{i<N} C_i) \leq \mu(B).$$

(ii) $\mu(B) \leq \sum_i \mu(B_i)$: By claim 2.

(iii) $B \in \Sigma$: Fix $\epsilon > 0$. Use the first part of the proof to find N with
$$\sum_{i<N} \mu(B_i) > \mu(B) - \epsilon/2.$$
Then find $C_i \subset B_i$ closed with
$$\mu(C_i) > \mu(B_i) - 2^{-i-1}\epsilon.$$
Then
$$\mu(\bigcup_{i<N} C_i) = \sum_{i<N} \mu(C_i) > (\sum_{i<N} \mu(B_i)) - \epsilon/2 > \mu(B) - \epsilon.$$

(Claim□)

Claim 7: If $B \in \Sigma$ then there is a closed C and open V with
$$C \subset B \subset V$$
and $\mu(V \setminus C) < \epsilon$.

Proof of Claim: Find C, V with $\mu(B) < \mu(C) + \epsilon/2$ and $\mu(B) > \mu(V) - \epsilon/2$. Claim 6 yields
$$\mu(V) = \mu(V \setminus C) + \mu(C),$$
thereby giving the required bound. (Claim□)

Claim 8: If $B_1, B_2 \in \Sigma$, then $B_1 \setminus B_2$ is in Σ.

Proof of Claim: Fix $\epsilon > 0$ and use claim 7 to choose $C_i \subset B_i \subset V_i$ with $\mu(V_i - C_i) < \epsilon/2$.
$$B_1 \setminus B_2 \subset (V_1 \setminus C_1) \cup (C_1 \setminus V_2) \cup (V_2 \setminus C_2).$$
Since these last three sets are disjoint, claim 2 yields
$$\mu(B_1 \setminus B_2) < \mu(C_1 \setminus V_2) + \epsilon.$$

(Claim□)

Claim 9: Σ is a Boolean algebra.

Proof of Claim:

Given A, B in Σ we can repeatedly claim 8 to
$$A \cap B = A \setminus (A \setminus B)$$
and claim 6 to
$$A \cup B = A \cup (B \setminus A).$$

(Claim□)

Claim 10: Σ is a σ-algebra.

Proof of Claim: Given $(B_i)_i$ all in Σ, let

$$C_i = B_i \setminus \bigcup_{j<i} B_j.$$

These are all in Σ by claim 9, and then by claim 6

$$\bigcup_i B_i = \bigcup_i C_i$$

is as well. (Claim□)

Σ is a σ-algebra incuding the closed sets, and thus every Borel set.

We have marched through vale and stormed through precarious mountain parts, but there is one last garrison of resistance. We still need to show the equality of integration by μ and application of Λ.

Claim 11: 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 it follows from claim 3. (Claim□)

Claim 12: For $f \in C(K)$ 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 5.2 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 11 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 13: For any $f \in C(K)$

$$\int f d\mu \leq \Lambda(f).$$

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 5.4 *If $\Lambda \in C(K)^*$ 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 $f_1, f_2, ..., f_n \in C(K)$.

Exercise (i) Let K be a compact metric space. Let $C(K, [-1, 1])$ be the subspace of $C(K)$ 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 $C(K)$ is the collection of linear, bounded, *real valued* functions.

Lemma 5.5 *Let $\Lambda \in C(K)^*$. Then there is a positive $\Phi \in C(K)^*$ 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□)

Then given any $f \in C(K)$ 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 5.6 *If $\Lambda \in C(K)^*$ then there is a signed measure ν with*

$$\Lambda(f) = \int f d\nu.$$

Proof Obtain Φ as in 5.5 and then let μ be as in 5.4 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 [6]) 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 5.7 *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 $f : K \to \mathbb{C}$.

6 Stone-Weierstrass

There is an extended discussion in [6] of this fundamental result, but I understand the proof was not actually presented in Math 312. I am going to present an alternate proof which makes use of the Riesz representation theorem.

As in the earlier theorems, I will always 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 $a = b = c$.

Definition Let $\mathbb{B}$ be a Banach space with dual space $\mathbb{B}^*$. The *weak star topology* on $\mathbb{B}^*$ is the one generated by the basic open sets

$$\{\varphi \in \mathbb{B}^* : \varphi(z_0) \in (a_0, b_0), \varphi(z_1) \in (a_1, b_1), ...\varphi(z_n) \in (a_n, b_n)\}$$

for $z_0, ..., z_n \in \mathbb{B}$, $a_0, ..., a_n, b_0, ..., b_n \in \mathbb{R}$. We say that a set is *weak star open* or *weak star closed* when it is open or, respectively, closed in this topology.

As a warning on notation, many people use "*weak**" rather than "weak star" to denote this topology.

Exercise (Alaoglu's theorem) Recall that 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.)

Exercise (i) Show that the basic open sets in the weak star topology are all convex.

(ii) Show that if $V \subset \mathbb{B}^*$ is convex and weak star open, then for any $\phi \in \mathbb{B}^*$ the set

$$V + \phi = \{\psi + \phi : \psi \in V\}$$

is also convex and weak star open as is

$$-V = \{-\psi : \psi \in V\}.$$

Lemma 6.1 *Let $\mathbb{B}$ be a Banach space and $A \subset \mathbb{B}^*$ convex and and weak star 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 star open neighborhood of the identity with $\psi + V \subset A$. Note that

$$t^{-1}(1-t)V =_{\mathrm{df}} \{t^{-1}(1-t)\varphi : \varphi \in V\}$$

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 6.2 *(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 star 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□)

Note that if $\psi \in V$, $\alpha \in (0,1)$ then the function

$$T : \phi \mapsto \alpha\psi + (1-\alpha)\phi$$

$$A \to A$$

is continuous, affine, injective, and has $T[\overline{V}] \subset V$ by 6.1. 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.

Claim: If W is a convex subset of A, then $V \cup W$ is a convex subset of A.

Proof of Claim: Since given any $\psi \in V$, $\phi \in W$, and $\alpha \in (0,1)$, the argument above showed that $\alpha\psi + (1-\alpha)\phi \in V$. (Claim□)

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. □

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 $|\mu| = \mu^+ + \mu^-$.

Exercise $\|\mu\| = \||\mu|\| = |\mu|(K)$.

(Hint: Here $\|\mu\|$ refers to its norm viewed as an element of $C(K)^*$, the dual space to $C(K)$. Since the continuous functions are dense in $L^1(\mu)$, this in turn equals the supremum over $\int f d\mu$ where $f : K \to [-1, 1]$ is measurable.)

Definition $\mathcal{A} \subset C(K)$ 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 6.3 *(Stone-Weierstrass) Let $\mathcal{A} \subset C(K)$ be an algebra which is closed in the sup norm, separates points, and contains the constant functions. Then $\mathcal{A} = C(K)$.*

Proof Consider the subset A of $C(K)^*$ 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. [6]) 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 6.2 to obtain some extreme $\Lambda \in A$, $\|\Lambda\| = 1$. Apply 5.3 to get a signed measure μ with

$$\Lambda(f) = \int f d\mu$$

all $f \in C(K)$. By the exercise above we have $|\mu|(K) = 1$. Let C be the *support* of $|\mu|$:

$$C =_{\text{df}} K \setminus \bigcup \{U : U \text{ open }, |\mu|(U) = 0\}.$$

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 $g \in C(K)$

$$\Lambda(g) = \Lambda(c_0),$$

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 $g \in C(K)$. (Claim□)

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 $f \in C(K)$ 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$\square$)

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. $\square$

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 5.7.

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 6.4 *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 $\mathcal{A} \subset C(K)$ is an algebra of functions which separates points, then $\mathcal{A}$ will be dense viewed as a subset of the Banach space $L^p(X, \mu)$.

7 Measure disintegration

One of the most important consequences of 5.3 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 5.4, $P(K)$ 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 7.1 *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 $\Sigma(A) = \{B \cap A : B \in \Sigma\}$.

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□)

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. □

Theorem 7.2 *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 7.3 *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 5.3, 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 1.6 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 7.4 *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 1.6 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 2.4 to find Borel functions $(f_C)_{C \text{ clopen}}$ with

$$\int_B f_C(y) d\nu(y) = \nu_C(B).^2$$

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

$$\int_B f_{\bigcup_i A_i}(y) d\nu(y) = \nu_{\bigcup_i A_i}(B)$$

$$= \sum_i \nu_{A_i}(B) = \sum \int_B f_{A_i}(y) d\mu$$

$$= \int_B \sum f_{A_i}(y) d\mu.$$

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

$$\mu^*_y : \{C_n : n \in \mathbb{N}\} \to [0, 1]$$

$$C_n \mapsto f_{C_n}(y)$$

[2]You may object. Literally as stated at 2.4, 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.[3]

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

$$\mu(A) = \int \mu_y(A) d\nu(y).$$

□

[3]But please do see me if you find yourself puzzled about what is going on.

8 Ergodic theory

8.1 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

$$\pi_\ell : x \mapsto e^{2\ell\pi ix} = \cos(2\ell\pi x) + i\sin(2\ell\pi x).$$

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.[4] One way to see this is to apply Stone-Weierstrass at 6.4 to conclude

[4]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 8.1 *(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 I toss a coin: If it comes up heads, I take a ball from the red urn and place it in the blue; if tails, I take one from the blue urn and place it in the red. (If at some stage one of the urns is empty and my coin instructs me to take a ball from that empty urn, then I just do nothing and flip again five seconds later.)

Intuitively we expect a mixing. Some balls should start to migrate across, and 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 [7]. 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.

8.2 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 [6]. 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

$$\forall g \in \mathcal{H}_0(\langle h, g\rangle = 0).$$

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 8.2 *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 8.3 *(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

$$\langle Uf, Ug\rangle = \langle f, g\rangle.$$

Note then that U will also be one to one ($\|Uf\| = \|f\|$ all $f \in \mathcal{H}$) and hence invertible.

Lemma 8.4 *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 8.5 *For T as above, U_T is a unitary operator.*

So much for the basics. Now for some true grit.

Lemma 8.6 *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

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k \bar{f} = \frac{1}{n}(Ug - U^{k+1}g) < \frac{\epsilon}{2}.$$

all $n > N$. Then

$$\|\frac{1}{n}\sum_{k=0}^{n-1} U^k f\| \leq \|\frac{1}{n}\sum_{k=0}^{n-1} U^k (f - \bar{f})\| + \|\frac{1}{n} Ug - U^{k+1} g\|$$

$$\leq \frac{1}{n}\sum_{k=0}^{n-1} \|f - \bar{f}\| + \frac{\epsilon}{2} < \epsilon$$

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 3.10 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 8.7 *(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 8.6 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 8.8 *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 8.9 *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 8.10 *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 $B_1, B_2, ...$ (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 $B_1 \cup ... \cup B_n$. 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

$$\int f_n d\mu = \frac{1}{N} \int \sum_{k=0}^{N-1} f_n \circ T^k d\mu$$

$$\geq \frac{-n}{N} \int \|f\|_1 d\mu,$$

which tends to zero as N heads towards infinity.

Corollary 8.11 *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 8.12 *(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 8.11 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 4.2 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 8.7.)

(iii) Since $g \le f$ we have at each n

$$\int \frac{1}{n}\sum_{k=0}^{n-1} g\circ T^k d\mu \le \int \frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k d\mu,$$

and hence

$$\gamma = \int \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} g\circ T^k d\mu \le \int \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} f\circ T^k d\mu = \alpha.$$

(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□)

Quantifying over all $\epsilon > 0$ we obtain

$$\alpha \ge \int f.$$

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

$$\alpha = \int \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k d\mu \leq \text{limsup}_{n\to\infty} \int \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k d\mu = \int f d\mu,$$

thereby completing the proof of the theorem. □

8.3 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 §8.1 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 §8.1. We let T act on $\mathbb{R}/\mathbb{Z}$ by some kind of irrational rotation: For instance

$$T(x) = x + \sqrt{2} \bmod 1.$$

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 [7] for a proof.)

8.4 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.[5]

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.

[5] 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.

9 The ergodic theory of general groups

9.1 Non-amenable groups

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 9.1 *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 3.6 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[6]. One of the deepest theorems in this entire area was proved in 2005:

Theorem 9.2 *(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 $\sigma \in F$.

[6]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.

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 and more can be found in [4].

Lemma 9.3 *$\mathbb{F}_2$ 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 9.4 *$\mathbb{Z}$ 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 9.5 *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 $A_i = \{x : xE_i\sigma \cdot x\}$. 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 9.6 *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.

9.2 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[7], 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),

$$E^{(0,1)} = \prod_E (0, 1),$$

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 9.7 *Let $\mathbb{F}_2 = \langle a, b\rangle$ and take as our generating set $S = \{a, b\}$. Then*

$$p_c(G(\mathbb{F}_2, S)) < 1.$$

[7]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 $\{w_i, w_{i+1}\} \in F$

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 9.5. □

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] 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.

[5] A. S. Kechris, **Classical Descriptive Set Theory,** Springer-Verlag, New-York, 1995.

[6] J. J. Koliha, **620-312 Linear Analysis,** unpublished notes.

[7] K. Petersen, **Ergodic Theory,** Cambridge University Press, Cambridge, 1983.

[8] W. Rudin, **Real and Complex Analysis,** Walter Rudin, McGraw-Hill, New York, 1966.