A VERY BRIEF REVIEW OF MEASURE THEORY

A brief philosophical discussion. Measure theory, as much as any branch of mathematics, is an area where it is important to be acquainted with the basic notions and statements, but not desperately important to be acquainted with the detailed proofs, which are often rather unilluminating. One should always have in a mind a place where one could go and look if one ever *did* need to understand a proof: for me, that place is Rudin's *Real and Complex Analysis* (Rudin's "red book").

If one wishes to do ergodic theory it is hopeless to try to pretend that measure and integration theory do not exist. It is vital to have access to all the limiting processes that are valid in measure theory but not permissible in weaker setups (such as, for example, the theory of Riemann integration).

In the course I deal exclusively with compact metric spaces. Measure theory in this setting most certainly does not convey all of the essence of the theory as a whole; for example our discussion excludes measure theory on $\mathbb{R}$. However it does allow for the avoidence of a few technicalities.

What is a measure? A measure is a way of assigning a *volume* to subsets of X. In most nontrivial settings one is not allowed to assign a volume to any old subset of X. The collection of sets that one is allowed to measure must be a σ-algebra: collections which contain $\emptyset$ and X and which are closed under complementation and taking countable unions. And, of course, if $\mathcal{F}$ is a σ-algebra the measure cannot be an arbitrary function on $\mathcal{F}$. In this course we will be dealing exclusively with *probability measures*, which are functions $\mu : \mathcal{F} \to [0,1]$ such that $\mu(X) = 1$ and $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ for any countable collection of *disjoint* sets $A_n \in \mathcal{F}$.

What can one do with a measure? By far the most important thing one can do with a measure is use it to integrate functions $f : X \to \mathbb{R}$. One cannot integrate arbitrary functions – f must be *measurable*, which means that each level set $f^{-1}([t,\infty))$ lies in the σ-algebra $\mathcal{F}$. As soon as one has made this definition, there are all sorts of things one might hope to prove, for example that the sum of two measurable functions or the composition of a measurable function with a continuous function is again measurable. All reasonable such statements are true but the proofs of these and subsequent straightforward facts occupy a lot of space.

The definition of the integral $\int f\,d\mu$ takes place in several stages. One first defines it for non-negative simple measurable functions, that is to say measurable functions $s : X \to \mathbb{R}_{\geqslant 0}$ with finite range. If $s(x) = \alpha_i$ for $x \in A_i$, $i = 1, \dots, k$, then we define (as is natural) $\int s\,d\mu := \sum_{i=1}^{k} \alpha_i \mu(A_i)$. If f is nonnegative, one then defines $\int f\,d\mu$ to be the supremum of all integrals $\int s\,d\mu$, over all simple measurable functions $s : X \to \mathbb{R}_{\geqslant 0}$ such that $0 \leqslant s \leqslant f$ pointwise. One can modify this in an obvious way for nonpositive functions f, and then define the integral for arbitrary measurable functions by splitting $f = f^+ + f^-$.

What properties does the integral enjoy? There is a huge list of obvious-seeming properties that the integral enjoys. For example, the map $f \mapsto \int f \, d\mu$ is linear. Perhaps the most important property, and really the *raison d'être* for the Lebesgue integral, is the following result, which states that integration and taking of limits are compatible.

Theorem 0.1 (Monotone convergence theorem)**.** *Suppose that* $(f_n)_{n=1}^{\infty}$ *are measurable functions and that* $0 \leqslant f_1(x) \leqslant f_2(x) \leqslant \dots$ *for all* x*. Suppose that* $f_n(x) \to f(x)$ *pointwise. Then* f *is measurable and* $\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu$*.*

Such a statement is definitely not true for the Riemann integral. Indeed if one takes $f_n(x)$ to be the function on $[0, 1]$ defined by $f_n(x) := 1$ if x is a rational with denominator $\leqslant n$ and $f_n(x) = 0$ otherwise then each f_n has Riemann integral zero. Furthermore we have $f_n \to f$ pointwise, where f is the characteristic function of $[0, 1] \cap \mathbb{Q}$. This limit function is not Riemann integrable.

Constructing interesting measures. The hardest results in measure theory, slightly embarrassingly, are required to construct interesting examples of measures. In this course the underlying set X will have the structure of a compact metric space with metric d, and we want to be able to compute the volume of the most obvious sets, namely the open balls $B(x, r) := \{y \in X : d(x, y) < r\}$. The smallest σ-algebra containing all the open balls is called the *Borel* σ*-algebra* and is generally denoted by $\mathcal{B}$. Elements of $\mathcal{B}$ are called Borel sets; amongst them are countable unions of closed sets (called F_δ-sets) and countable intersections of open sets (called G_σ-sets). As we showed earlier in the course, X is separable. This means that $\mathcal{B}$ contains the open sets of X (in fact this is usually the definition of $\mathcal{B}$).

Theorem 0.2 (Lebesgue measure on $\mathbb{R}/\mathbb{Z}$)**.** *There is a unique probability measure* μ *on* $\mathbb{R}/\mathbb{Z}$ *such that* $\mu((a, b)) = b - a$ *for all* $0 \leqslant a \leqslant b \leqslant 1$*.*

The existence of Lebesgue measure is a special case of a much more general theorem, the Riesz Representation theorem.

Theorem 0.3 (Riesz representation theorem)**.** *Let* X *be a compact metric space Then the probability measures on the Borel* σ*-algebra* $\mathcal{B}$ *are in one-to-one correspondence with positive linear functionals* $\Lambda : C(X) \to \mathbb{R}$ *with* $\Lambda 1 = 1$ *via the correspondence* $\Lambda f \leftrightarrow \int f \, d\mu$*.*

Here, $C(X)$ is the space of continuous functions from X to $\mathbb{R}$, and by a positive linear functional we mean a linear functional with the additional property that $\Lambda f \geqslant 0$ whenever $f \geqslant 0$ pointwise.

It is not actually a trivial matter to deduce the existence of Lebesgue measure from this theorem. In fact that linear functional Λ that is appropriate here is none other than the Riemann integral, which is of course well-defined and linear when restricted to continuous functions.

In fact the proof of the Riesz representation theorem and a little extra work shows that any probability measure on $\mathcal{B}$ is automatically *regular*, which means that

$$\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ closed}\} = \inf\{\mu(U) : A \subseteq U, U \text{ open}\}.$$

For the purposes of this course it might be better to regard regularity as patr of the definition of a measure: we will not encounter measures which are not regular.

The Riesz representation theorem has two directions. The statement that every probability measure μ gives rise to a linear functional Λ is relatively straightforward, and is essentially just the construction of the integral $\int f\, d\mu$. The proof of the other direction is rather long; this is where the measure μ actually gets constructed. The key idea is to use the notion of regularity to *define* μ. One first defines μ on open sets (and hence, by complementation, on closed sets) by setting $\mu(U) := \sup\{\Lambda f : f \in C(X), 0 \leqslant f \leqslant 1_U$. One then defines μ^+ of an arbitrary subset $A \subseteq X$ to be $\inf\{\mu(U) : A \subseteq U, U \text{ open}\}$ and $\mu^-(A)$ to be $\sup\{\mu(K) : K \subseteq A, K \text{ closed}\}$. One then shows that the collection of sets $A \subseteq X$ for which $\mu^-(A) = \mu^+(A)$ is a σ-algebra $\tilde{\mathcal{B}}$ containing $\mathcal{B}$, and that $\mu = \mu^+ = \mu^-$ is a measure on $\tilde{\mathcal{B}}$ with the property that $\Lambda f = \int f\, d\mu$ for $f \in C(X)$. There is much to be checked: it is not an easy theorem.

There is another important respect in which the Riesz representation theorem gives something stronger than we have stated. The measure μ is actually a measure on a σ-algebra $\tilde{\mathcal{B}}$ which, in general, is strictly larger than the Borel σ-algebra $\mathcal{B}$. It has the additional property that if $A \in \tilde{\mathcal{B}}$ and $\mu(A) = 0$ then any subset of A is also in $\tilde{\mathcal{B}}$. This property need not hold for $\mathcal{B}$ itself: when $X = [-1, 1]$ for example one may employ a cardinality argument to establish that there are subsets of the Cantor set on $[-1, 1]$ which do not lie in $\mathcal{B}$. This is a technical point and we will not need to dwell on it in the course.

Limits of measures. The Riesz representation theorem tells us that probability measures on X are in 1-1 correspondence with elements $\Lambda \in C(X)^*$ which are positive and normalised so that $\Lambda 1 = 1$. This is, in particular, a convex subset of $C(X)^*$. Write $\mathcal{M}(X)$ for the space of regular Borel measures on X. The identification of $\mathcal{M}(X)$ with a convex subset of $C(X)^*$ makes it much easier to study the former object, and in particular to discuss limits of measures.

Definition 0.4 (Weak convergence of measures)**.** Let μ and μ_n, $n = 1, 2, \ldots$, be measures in $\mathcal{M}(X)$. Then we say that μ_n converges weakly to μ, and write $\mu_n \to \mu$, if, for every $f \in C(X)$, we have $\int f\, d\mu_n \to \int f d\mu$.

An important fact is that $\mathcal{M}(X)$ is compact in the topology of weak convergence.

Proposition 0.5 (Space of probability measures is weakly compact)**.** *Suppose that $(\mu_n)_{n=1}^\infty$ is a sequence of probability measures. Then there is a subsequence $(\mu_{n_k})_{k=1}^\infty$ which converges weakly to some $\mu \in \mathcal{M}(X)$.*

Proof. By the Riesz representation theorem it suffices to prove that the closed unit ball of $C(X)^*$ is compact in the weak topology: that is, if $\Lambda_n : X \to \mathbb{R}$ are functionals with $\|\Lambda_n\| \leqslant 1$ then there is some subsequence $(\Lambda_{n_k})_{k=1}^\infty$ which tends weakly to a functional Λ in the sense that $\Lambda_n f \to \Lambda f$ for all $f \in C(X)$.

Note that if each Λ_n corresponds to a probability measure μ_n then we have $|\Lambda_n f| \leqslant \|f\|_\infty \int \mu_n = \|f\|_\infty$, and so Λ_n does lie in the unit ball of $C(X)^*$. It is clear that if all the Λ_n are positive and normalised so that $\Lambda_n 1 = 1$ then the same will be true for any weak limit Λ; thus by the Riesz representation theorem such a limit corresponds to a probability measure μ.

The statement that the closed unit ball of $C(X)^*$ is compact in the weak topology is known as the *Banach-Alaoglu theorem*, and it is usually proved via Tychonov's theorem. In our setting, where X is a compact metric space, a more direct and vaguely constructive proof using a diagonalisation argument is possible. One might compare this with the rather simpler diagonal argument we used to prove that $\Lambda^{\mathbb{Z}}$ is sequentially compact.

We begin by recalling that $C(X)$ is separable (has a countable dense subset). This follows from a version of the Stone-Weierstrass theorem.

Take, then, a countable dense collection of functions $f_1, f_2, \dots$ in $C(X)$. Consider the sequence $\Lambda_1 f_1, \Lambda_2 f_1, \dots$. We may find a subsequence $(n_{1,i})_{i=1}^{\infty}$ of $\mathbb{N}$ such that the sequence $(\Lambda_{n_{1,i}} f_1)_{i=1}^{\infty}$ converges. We may then pass to a further subsequence $(n_{2,i})_{i=1}^{\infty}$ such that the sequence $(\Lambda_{n_{2,i}} f_2)_{i=1}^{\infty}$ converges, and so on. Set $n_i := n_{i,i}$. Then the diagonal sequence $(n_i)_{i=1}^{\infty}$ has the property that $(\Lambda_{n_i} f_k)_{i=1}^{\infty}$ converges for all k. Define $\Lambda f_k := \lim_{i\to\infty} \Lambda_{n_i} f_k$ for $k = 1, 2, \dots$. We extend this to a map $\Lambda : C(X) \to \mathbb{R}$ by defining $\Lambda f := \lim_{j\to\infty} \Lambda f_{k_j}$, for any sequence $(k_j)_{j=1}^{\infty}$ such that $f_{k_j} \to f$ in $C(X)$.

We claim that $\Lambda \in B_1(C(X)^*)$ and that $\Lambda_{n_i} \to \Lambda$ in the weak topology. There is much to prove here (for example we have not yet shown that Λ is well-defined, less still that it is a bounded linear functional). This is a somewhat tedious task which we leave to the reader. □

L^p-spaces and L^p-norms. One of the most important things that measure theory allows us to do is to define these spaces of functions. Let X be a compact metric space with a regular Borel probability measure μ. If $f : X \to \mathbb{C}$ is a function we define

$$\|f\|_p := \Big(\int |f|^p d\,\mu \Big)^{1/p}$$

for $1 \leqslant p < \infty$ and $\|f\|_\infty$ to be the essential supremum of f, that is to say the infimum of all those numbers M such that $|f(x)| \leqslant M$ outside of a set of measure zero (*almost everywhere*).

These objects $\|\cdot\|_p$ satisfy the triangle inequality $\|f + g\|_p \leqslant \|_p + \|g\|_p$ and also the inequality $\|\lambda f\|_p \leqslant |\lambda|\|f\|_p$ for complex scalars λ. This qualifies them as *seminorms*; they are not fully-fledged norms because it is possible to have $\|f\|_p = 0$ without f being zero. This is the case if, and only if, f vanishes almost everywhere. We write $L^p(X)$ for the space of all measurable functions $f : X \to \mathbb{C}$ with $\|f\|_p < \infty$, quotiented out by the equivalence relation of being "equal almost everywhere". One does not introduce any special notation for these equivalence classes, abusing notation by regarding $L^p(X)$ as a space of functions.

The L^p-norms are nested: $\|f\|_p \leqslant \|f\|_{p'}$ whenever $1 \leqslant p \leqslant p' \leqslant \infty$. This follows from Hölder's inequality and it is important here that μ is a probability measure, that is to say $\int 1\, d\mu = 1$.

A very important fact about the $L^p(X)$ spaces is that they are complete (and hence each one is a Banach space). The truth of this statement is a very important justification for introducing measures.

Approximation of measurable functions. In practice one rarely tries to understand anything about general measurable functions or even functions in $L^p(X)$. Instead one approaches them by stealth, as limits of functions which are much easier to understand.

As we are in Cambridge, it seems appropriate at this point to mention J. E. Littlewood's three basic principles. These can be a useful practical guide to working with measures; it is notable that they were formulated at a time when the use of rough heuristics and models to motivate quite technical subjects was not nearly so widespread as it is now.

Here are the three principles, which apply to any regular Borel measure μ on a compact metric space X:

(i) A measurable set is nearly an open set;
(ii) A measurable function is nearly a continuous function;
(iii) A convergent sequence of functions is nearly uniformly convergent.

Let us briefly discuss the three principles in turn. A precise version of the first follows immediately from the definition of a regular measure, viz that $\mu(E) = \inf_{E \subseteq U} \mu(U)$ for any measurable set E, the infimum being taken over all open sets U containing E. Thus for any $\varepsilon > 0$, there is an open set U whose symmetric difference with E has measure less than ε.

The second point refers to *Lusin's theorem.* If f is any measurable function and if $\varepsilon > 0$, this states that there is a continuous function $g \in C(X)$ such that $f(x) = g(x)$ except on a set of measure at most ε. Note that it does *not* assert that f is continuous except on a set of measure ε.

The third point refers to Egorov's theorem: If (f_n) is a sequence of measurable functions which converge pointwise on X, and if $\varepsilon > 0$, then there is a measurable set $X' \subseteq X$, $\mu(X \setminus X') \leqslant \varepsilon$, such that the f_n converge uniformly on X'.

Perhaps the most useful application of these principles is the following straightforward consequence of the second one: the space of continuous functions $C(X)$ is dense in $L^p(X)$, for all $1 \leqslant p \leqslant \infty$.

When the space X has additional structure, one can often pass to an even nicer set of functions which is dense in $L^p(X)$. If X is a smooth manifold, for example, the space $C^\infty(X)$ of smooth functions is dense. If $X = \mathbb{R}^d/\mathbb{Z}^d$ is a torus then the trigonometric polynomials $\sum_{|r| \leqslant R} a_r e^{2\pi i r \cdot \theta}$ are dense (see the example sheet).

Some philosophical remarks by Akshay Venkatesh. I hope that the pleasant properties of measures presented here, particularly their closure properties under taking limits, will convince you that they are the "right" objects to consider. Here are some further remarks, by Akshay Venkatesh, which will make more sense a little later in the course.

What is gained by going through measures? Measures have much better formal properties than sets. A particularly important difference is that a T-invariant probability measure can be decomposed into "minimal" invariant measures (the ergodic decomposition). That property does not seem to have a clean analogy at the level of T-invariant closed sets. In particular, although a T-invariant closed set always contains a minimal T-invariant closed set, it cannot be decomposed into such sets in any obvious way.

Further reading. I rather like the introduction to Lebesgue measure on $\mathbb{R}$ in the book of Stein and Shakarchi, *Real analysis: measure theory, integration and Hilbert spaces.* For a comprehensive introduction to the more general setting that we need here one might consult Rudin's "red" book, *Real and complex analysis.* He works with locally compact Hausdorff spaces X rather than simply compact metric spaces as we discuss here. The words of Akshay Venkatesh are taken from his article *The work of Einsiedler, Katok and Lindenstrauss on the Littlewood conjecture,* Bull. Amer. Math. Soc. **45** (2008), 117-134.