8. Measure spaces

§8.1 Background

In Lecture 1 we remarked that ergodic theory is the study of the qualitative distributional properties of typical orbits of a dynamical system and that these properties are expressed in terms of measure theory. Measure theory therefore lies at the heart of ergodic theory. However, we will not need to know the (many!) intricacies of measure theory and the next few lectures will be devoted to an expository account of the required facts.

§8.2 Measure spaces

Loosely speaking, a measure is a function that, when given a subset of a space X, will say how 'big' that subset is. A motivating example is given by Lebesgue measure. The Lebesgue measure of an interval is given by its length. In defining an abstract measure space, we will be taking the properties of 'length' (or, in higher dimensions, 'volume') and abstracting them, in much the same way that a metric space abstracts the properties of 'distance'.

It turns out that in general it is not possible to be able to define the measure of an arbitrary subset of X. Instead, we will usually have to restrict our attention to a class of subsets of X.

Definition. A collection $\mathcal{B}$ of subsets of X is called a *σ-algebra* if

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

(ii) if $E \in \mathcal{B}$ then its complement $X \setminus E \in \mathcal{B}$,

(iii) if $E_n \in \mathcal{B}$, $n = 1, 2, 3, \ldots$, is a countable sequence of sets in $\mathcal{B}$ then their union $\bigcup_{n=1}^{\infty} E_n \in \mathcal{B}$.

Examples.

1. The trivial σ-algebra is given by $\mathcal{B} = \{\emptyset, X\}$.
2. The full σ-algebra is given by $\mathcal{B} = \mathcal{P}(X)$, i.e. the collection of *all* subsets of X.

Here are some easy properties of σ-algebras:

Lemma 8.1
Let $\mathcal{B}$ be a σ-algebra of subsets of X. Then

(i) $X \in \mathcal{B}$*;*

(ii) if $E_n \in \mathcal{B}$ *then* $\bigcap_{n=1}^{\infty} E_n \in \mathcal{B}$*.*

Exercise 8.1
Prove Lemma 8.1.

In the special case when X is a compact metric space there is a particularly important σ-algebra.

Definition. Let X be a compact metric space. We define the *Borel σ-algebra* $\mathcal{B}(X)$ to be the smallest σ-algebra of subsets of X which contains all the open subsets of X.

Remarks.

1. By 'smallest' we mean that if $\mathcal{C}$ is another σ-algebra that contains all open subsets of X then $\mathcal{B}(X) \subset \mathcal{C}$.
2. We say that the Borel σ-algebra is generated by the open sets. We call sets in $\mathcal{B}(X)$ a *Borel set*.
3. By Definition 8.2(ii), the Borel σ-algebra also contains all closed sets and is the smallest σ-algebra with this property.

Let X be a set and let $\mathcal{B}$ be a σ-algebra of subsets of X.

Definition. A function $\mu : \mathcal{B} \to \mathbb{R}^+ \cup \{\infty\}$ is called a *measure* if:

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

(ii) if E_n is a countable collection of pairwise disjoint sets in $\mathcal{B}$ (i.e. $E_n \cap E_m = \emptyset$ for $n \neq m$) then

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n).$$

(If $\mu(X) < \infty$ then we call μ a *finite measure*.) We call $(X, \mathcal{B}, \mu)$ a *measure space*.

If $\mu(X) = 1$ then we call μ a *probability* or *probability measure* and refer to $(X, \mathcal{B}, \mu)$ as a *probability space*.

Remark. Thus a measure just abstracts properties of 'length' or 'volume'. Condition (i) says that the empty set has zero length, and condition (ii) says that the length of a disjoint union is the sum of the lengths of the individual sets.

Definition. We say that a property holds *almost everywhere* if the set of points on which the property fails to hold has measure zero.

We will usually be interested in studying measures on the Borel σ-algebra of a compact metric space X. To define such a measure, we need to define the measure of an arbitrary Borel set. In general, the Borel σ-algebra is extremely large. In the next section we see that it is often unnecessary to do this and instead it is sufficient to define the measure of a certain class of subsets.

§8.3 The Kolmogorov Extension Theorem

A collection $\mathcal{A}$ of subsets of X is called an *algebra* if:

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

(ii) if $A, B \in \mathcal{A}$ then $A \cap B \in \mathcal{A}$;

(iii) if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.

Thus an algebra is like a σ-algebra, except that we do not assume that $\mathcal{A}$ is closed under countable unions.

Example. Take $X = [0, 1]$, and $\mathcal{A} = \{\text{all finite unions of subintervals}\}$.

Let $\mathcal{B}(\mathcal{A})$ denote the σ-algebra generated by $\mathcal{A}$, i.e., the smallest σ-algebra containing $\mathcal{A}$. (In the above example $\mathcal{B}(\mathcal{A})$ is the Borel σ-algebra.)

Theorem 8.2 (Kolmogorov Extension Theorem)
Let $\mathcal{A}$ be an algebra of subsets of X. Suppose that $\mu : \mathcal{A} \to \mathbb{R}^+$ satisfies:

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

(ii) there exists finitely or countably many sets $X_n \in \mathcal{A}$ such that $X = \bigcup_n X_n$ and $\mu(X_n) < \infty$;

(iii) if $E_n \in \mathcal{A}$, $n \geq 1$, are pairwise disjoint and if $\bigcup_{n=1}^\infty E_n \in \mathcal{A}$ then

$$\mu\left(\bigcup_{n=1}^\infty E_n\right) = \sum_{n=1}^\infty \mu(E_n).$$

Then there is a unique measure $\mu : \mathcal{B}(\mathcal{A}) \to \mathbb{R}^+$ which is an extension of $\mu : \mathcal{A} \to \mathbb{R}^+$.

Remarks.

(i) The important hypotheses are (i) and (iii). Thus the Kolmogorov Extension Theorem says that if we have a function μ that looks like a measure on an algebra $\mathcal{A}$, then it is indeed a measure when extended to $\mathcal{B}(\mathcal{A})$.

(ii) We will often use the Kolmogorov Extension Theorem as follows. Take $X = [0, 1]$ and take $\mathcal{A}$ to be the algebra consisting of all finite unions of sub-intervals of X. We then define the 'measure' μ of a subinterval in such a way as to be consistent with the hypotheses of the Kolmogorov Extension Theorem. It then follows that μ does indeed define a measure on the Borel σ-algebra.

(iii) Here is another way in which we shall use the Kolmogorov Extension Theorem. Suppose we have two measures, μ and ν, and we want to see if $\mu = \nu$. A priori we would have to check that $\mu(B) = \nu(B)$ for all $B \in \mathcal{B}$. The Kolmogorov Extension Theorem says that it is sufficient to check that $\mu(E) = \nu(E)$ for all E in an algebra $\mathcal{A}$ that generates $\mathcal{B}$. For example, to show that two measures on $[0, 1]$ are equal, it is sufficient to show that they give the same measure to each subinterval.

§8.4 Examples of measure spaces

Lebesgue measure on $[0, 1]$. Take $X = [0, 1]$ and take $\mathcal{A}$ to be the collection of all finite unions of sub-intervals of $[0, 1]$. For a sub-interval $[a, b]$ define

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

This satisfies the hypotheses of the Kolmogorov Extension Theorem, and so defines a measure on the Borel σ-algebra $\mathcal{B}$. This is *Lebesgue measure*.

Lebesgue measure on $\mathbb{R}/\mathbb{Z}$. Take $X = \mathbb{R}/\mathbb{Z} = [0, 1) \bmod 1$ and take $\mathcal{A}$ to be the collection of all finite unions of sub-intervals of $[0, 1)$. For a sub-interval $[a, b]$ define

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

This satisfies the hypotheses of the Kolmogorov Extension Theorem, and so defines a measure on the Borel σ-algebra $\mathcal{B}$. This is *Lebesgue measure on the circle*.

Lebesgue measure on the k-dimensional torus. Take $X = \mathbb{R}^k/\mathbb{Z}^k = [0, 1)^k \bmod 1$ and take $\mathcal{A}$ to be the collection of all finite unions of k-dimensional sub-cubes $\prod_{j=1}^k [a_j, b_j]$ of $[0, 1)^k$. For a sub-cube $\prod_{j=1}^k [a_j, b_j]$ of $[0, 1)^k$, define

$$\mu\left(\prod_{j=1}^k [a_j, b_j]\right) = \prod_{j=1}^k (b_j - a_j).$$

This satisfies the hypotheses of the Kolmogorov Extension Theorem, and so defines a measure on the Borel σ-algebra $\mathcal{B}$. This is *Lebesgue measure on the torus*.

Stieltjes measures. Take $X = [0,1]$ and let $\rho : [0,1] \to \mathbb{R}^+$ be an increasing function such that $\rho(1) - \rho(0) = 1$. Take $\mathcal{A}$ to be the algebra of finite unions of subintervals and define

$$\mu_\rho([a,b]) = \rho(b) - \rho(a).$$

This satisfies the hypotheses of the Kolmogorov Extension Theorem, and so defines a measure on the Borel σ-algebra $\mathcal{B}$. We say that μ_ρ is the measure on $[0,1]$ with density ρ.

Dirac measures. Finally, we give an example of a class of measures that do not fall into the above categories. Let X be an arbitrary space and let $\mathcal{B}$ be an arbitrary σ-algebra. Let $x \in X$. Define the measure δ_x by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

Then δ_x defines a probability measure. It is called the *Dirac measure at x*.

§8.5 Appendix: The Riemann integral

(This subsection is included for general interest and is not examinable!)

You have probably seen the construction of the Riemann integral. This gives a method for defining the integral of functions defined on $[a,b]$. In the next lecture we will see how the Lebesgue integral is a generalisation of the Riemann integral, in the sense that it allows us to integrate functions defined on spaces more general than subintervals of $\mathbb{R}$. However, the Lebesgue integral has other nice properties, for example it is well-behaved with respect to limits. Here we give a brief exposition about some inadequacies of the Riemann integral and how they motivate the Lebesgue integral.

Let $f : [a,b] \to \mathbb{R}$ be a bounded function (for the moment we impose no other conditions on f).

A *partition* Δ of $[a,b]$ is a finite set of points $\Delta = \{x_0, x_1, x_2, \ldots, x_n\}$ with

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b.$$

In other words, we are dividing $[a,b]$ up into subintervals.

We then form the *upper* and *lower Riemann sums*

$$\begin{aligned} U(f,\Delta) &= \sum_{i=0}^{n-1} \sup_{x \in [x_i, x_{i+1}]} f(x)\,(x_{i+1} - x_i), \\ L(f,\Delta) &= \sum_{i=0}^{n-1} \inf_{x \in [x_i, x_{i+1}]} f(x)\,(x_{i+1} - x_i). \end{aligned}$$

The idea is then that if we make the subintervals in the partition small, these sums will be a good approximation to (our intuitive notion of) the integral of f over $[a, b]$. More precisely, if

$$\inf_{\Delta} U(f, \Delta) = \sup_{\Delta} L(f, \Delta),$$

where the infimum and supremum are taken over all possible partitions of $[a, b]$, then we write

$$\int_a^b f(x)\, dx$$

for their common value and call it the *(Riemann) integral* of f between those limits. We also say that f is *Riemann integrable*.

The class of Riemann integrable functions includes continuous functions and step functions (i.e. finite linear combinations of characteristic functions of intervals).

However, there are many functions for which one wishes to define an integral but which are not Riemann integrable, making the theory rather unsatisfactory. For example, define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{otherwise.} \end{cases}$$

Since between any two distinct real numbers we can find both a rational number and an irrational number, given $0 \leq y < z \leq 1$, we can find $y < x < z$ with $f(x) = 1$ and $y < x' < z$ with $f(x') = 0$. Hence for any partition $\Delta = \{x_0, x_1, \ldots, x_n\}$ of $[0, 1]$, we have

$$\begin{aligned} U(f, \Delta) &= \sum_{i=0}^{n-1} (x_{i+1} - x_i) = 1, \\ L(f, \Delta) &= 0. \end{aligned}$$

Taking the infimum and supremum, respectively, over all partitions Δ, shows that f is *not* Riemann integrable.

Why does Riemann integration not work for the above function and how could we go about improving it? Let us look again at (and slightly rewrite) the formulæ for $U(f, \Delta)$ and $L(f, \Delta)$. We have

$$U(f, \Delta) = \sum_{i=0}^{n-1} \sup_{x \in [x_i, x_{i+1}]} f(x)\ l([x_i, x_{i+1}])$$

and

$$L(f, \Delta) = \sum_{i=0}^{n-1} \inf_{x \in [x_i, x_{i+1}]} f(x)\ l([x_i, x_{i+1}]),$$

where, for an interval $[y, z]$,

$$l([y,z]) = z - y$$

denotes its *length*. In the example above, things didn't work because dividing $[0, 1]$ into intervals (no matter how small) did not 'separate out' the different values that f could take. But suppose we had a notion of 'length' that worked for more general sets than intervals. Then we could do better by considering more complicated 'partitions' of $[0, 1]$, where by partition we now mean a collection of subsets $\{E_1, \ldots, E_m\}$ of $[0, 1]$ such that $E_i \cap E_j = \emptyset$, if $i \neq j$, and $\bigcup_{i=1}^{m} E_i = [0, 1]$.

In the example, for instance, it might be reasonable to write

$$\begin{aligned} \int_0^1 f(x)\,dx &= 1 \times l([0,1] \cap \mathbb{Q}) + 0 \times l([0,1] \backslash \mathbb{Q}) \\ &= l([0,1] \cap \mathbb{Q}). \end{aligned}$$

Instead of using subintervals, the Lebesgue integral uses a much wider class of subsets (namely sets in the given σ-algebra) together with a notion of 'generalised length' (namely, measure).