Measure Theory

Jacob Denson

October 25, 2017

Table Of Contents

Chapter 1

The Lebesgue Measure

Area is one of the most primitive measurements in geometry. Every elementary school student knows that the area of a circle of radius r is πr^2, and that the area of a rectangle is equal to the product of the lengths of its two distinct side lengths. But given a general shape in the plane, it suddenly becomes very difficult to determine a shape's area. In the work of the ancient Greeks, especially Archimedes, we find two methods of finding more complex areas:

- If a shape can be cut up into finitely many components, and then rearranged into the form of a different shape by a series of rigid motions, then the new shape has the same area as the old.
- We can obtain upper bounds on the area of the shape by enveloping the shape by another shape with an already known area, and lower bounds by finding shapes enclosed by the shape.

But with the advent of Cartesian coordinates and the subsequent modern introduction of set theory, shapes are identified with subsets of $\mathbf{R}^2$, and it is no longer clear how to define the area of a general subset of $\mathbf{R}^2$. More generally, it isn't clear how to define the lengths of subsets of the real line $\mathbf{R}$, or volumes of general subsets of $\mathbf{R}^3$. These are all *measures* of size in their relative dimensions, and so we call the general study of these obects *measure theory*. In line with this unification, we will let $|S|$, or $\mu(S)$ denote the length, area, and volume of a given shape S.

You might argue that these paradoxes are irrelevant in terms of any subset that occurs in modern mathematics, but this is not so. For instance, one often wants to measure the length of continuous curves in the plane, which can be considered another measure on space. In 1890, Peano constructed a continuous curve which covers the entire plane. The length of this curve is certainly suspect, because it is constructed as the limit of piecewise differentiable curves whose lengths tend to infinity. Modern developments even brought the ancient methods of area and volume into question. In 1924, Stefan Banach and Alfred Tarski showed that one can decompose a sphere into a finite number of components, which by a number of rigid motions can be rearranged into two copies of the original sphere! We seem to have produced two equal things from one thing – a feat not far from biblical miracle (though we have to use oranges, not loaves and fishes). In duplicating the sphere, Banach and Tarski showed that even the old methods of area geometry do not stand up to the techniques of modern mathematics, and a reanalysis of the entire field was necessary.

One key idea of measure theory is that the old methods of geometry continue to work, provided that we only concentrate on certain 'nice' subsets of space which obey our intuitions about size. Stefan Banach and Alfred Tarski engineered one of the first partitions of space which do not obey geometric intuition – we call these **unmeasurable sets**, because trying to measure their size causes problems. The main theorems of measure theory only work when we work with **measurable sets**. Indeed, we can reinterpret the methods of ancient geometry into two principles of measure theory:

- If a *measurable* set S has a decomposition into disjoint, *measurable* subsets $S_1, \dots, S_n$, then $|S| = \sum |S_i|$, and if S is transformed into T by a rigid motion, then $|S| = |T|$.

- If a set S is *measurable*, then there is a decreasing sequence of shapes $S_1, S_2, \dots$, each containing S, and each decomposable into disjoint intervals/squares/boxes (which we can easily measure the area of), with $|S_i| \to |S|$.

We will begin by extending the notion of sets to fairly general subsets of $\mathbf{R}^n$. We do this not only because this is classically how measure theory was introduced, but also because it brings to light the many intricate parts of the theory which we consider when we build measures on more general 'measure spaces'. We remind the reader that we shall use μ for the length of a subset of $\mathbf{R}$, the area of a subset of $\mathbf{R}^2$, the volume of a subset of $\mathbf{R}^3$, and higher dimensional variants.

1.1 Measuring Elementary Sets

Let's begin by using basic ideas of Euclidean geometry to find a basic class of sets which we can measure the size of without introducing paradoxes. The length of an **interval** with start point a and end point b, either closed, open, or half open, is $b - a$. We know from elementary geometry that the area of a rectangle is the product of the length of the intervals that define it, and we can generalize this to defining the measure of a general box, formed from the product of intervals. That is, if $I = I_1 \dots I_n$ is a box, where I_i is an interval starting at a_i and ending at b_i, then

$$|I| = (b_1 - a_1)\dots(b_n - a_n) = |I_1|\dots|I_n|$$

In general, the easiest sets to measure the area of are those covered by boxes, and we will show this leads to a system of area with a well defined theory.

Lemma 1.1. *If we decompose a box R into the disjoint union of finitely many boxes R_i, then the measure of the box is the sum of the measures of the boxes in the decomposition, in the sense that $\mu(R) = \sum \mu(R_i)$*

Proof. We proceed by a grid decomposition. Suppose first that the rectangular decomposition forms a grid, in the sense that we can index the decomposition as $R_{i_1 \dots i_n}$, where $R_{i_1 \dots i_n} = I^1_{i_1} \times \dots \times I^n_{i_n}$, and the endpoint of $I^k_{i_n}$ is the startpoint of $I^k_{i_n+1}$. Then

$$\begin{aligned}\sum_{i_1,\dots,i_n} \mu(R_{i_1 \dots i_n}) &= \sum_{i_1,\dots,i_n} \mu(I^1_{i_1})\dots\mu(I^n_{i_n}) \\ &= \prod_{k=1}^n \sum_j \mu(I^k_j)\end{aligned}$$

and the theorem is implied in this case by showing that $\sum_j \mu(I^k_j) = \mu(I_k)$, where $R = I_1 \times \dots \times I_n$. But this follows because the sum $\sum_j \mu(I^k_j)$ is a telescoping sum, with the highest indexed interval's endpoint equal to the endpoint of I_k, and the lower indices startpoint equal to the startpoint of

I_k. In general, it suffices to break a general decomposition into a further decomposition forming a grid, in such a way that the sum of the boxes in the first decomposition is equal to the sum of the boxes in the second. This is proven by forming the grid in each dimension, applying another telescoping sum type argument along each dimension. □

A similar grid decomposition like argument proves the following.

Lemma 1.2. *If a family of boxes $R_1, \dots, R_n$ covers R, then $\mu(R) \leqslant \sum \mu(R_i)$.*

Lemma 1.3. *If $R_1, \dots, R_n$ and $S_1, \dots, S_m$ are two disjoint families of boxess with $\bigcup R_i = \bigcup S_i$, then $\sum \mu(R_i) = \sum \mu(S_i)$.*

Alternatively, these theorems can be shown using a discretization argument. We begin by showing that

$$\mu(I) = \lim_{N \to \infty} \frac{|\mathbf{Z}/N \cap I|}{N}$$

From this, it is easy to argue that for any box R,

$$\mu(R) = \lim_{N \to \infty} \frac{|\mathbf{Z}^n/N \cap R|}{N^n}$$

But now if we write $R = \cup R_i$ as the union of disjoint intervals, then

$$\mu(R) = \lim_{N \to \infty} \frac{|\mathbf{Z}^n/N \cap R|}{N^n} = \sum_i \lim_{N \to \infty} \frac{|\mathbf{Z}^n/N \cap R_i|}{N^n} = \sum_i \mu(R_i)$$

and this proves the theorem. One might be tempted to define the measure of an arbitrary subset of $\mathbf{R}^n$ by the formula

$$\mu(E) = \lim_{N \to \infty} \frac{|\mathbf{Z}^n/N \cap E|}{N^n}$$

however, this definition runs into problems. One can find sets where this limit doesn't exist, and even if the limit does exist, we might not even have translation invariance. For instance, with respect to this function $\mathbf{Q} \cap [0,1]$ has length 1, but $\mathbf{Q} + \sqrt{2}$ has length 0. The definition is valid for all *Jordan measurable sets*. A more suitable way to obtain a continuous measure from some kind of discrete measure is by the theory of Monte Carlo integration, which we won't cover here.

The above lemmas guarantee that if $E \subset \mathbf{R}^n$ is the disjoint union of boxes $R_1, \dots, R_n$, then the definitions

$$\mu(E) = \sum \mu(R_i)$$

is well defined. We call a set like E an **elementary set**. We shall find that the *algebraic structure* of the family of sets a measure is defined over is interesting. One often has to consider the areas of the set formed from the union of two sets, or the intersection, and it is useful to know that we can measure the size of a set if it is the union of measurable sets, or the intersection of measurable sets. The next lemma is very useful in that regard

Lemma 1.4. *If R is the finite union of boxes, then R is the finite union of disjoint boxes.*

Proof. One can only prove this by a grid decomposition argument. □

Since the union of two sets which are finite unions of boxes is also a finite union of boxes, we conclude that the union of two elementary sets is also elementary.

Lemma 1.5. *The intersection of two box is a box.*

Proof. Note first that if I is an interval with start point a and endpoint b, and J is an interval with start point c and endpoint d, then $I \cap J$ is either empty, or an interval with startpoint $\max(a,c)$, and endpoint $\min(b,d)$. But then give $R = I_1 \times \dots \times I_n$ and $S = J_1 \times \dots \times J_n$, then

$$R \cap S = (I_1 \cap J_1) \times \dots \times (I_n \cap J_n)$$

which is a box. □

If $E = \bigcup R_i$ and $F = \bigcup S_i$ are elementary sets, then $E \cap F = \bigcup (R_i \cap S_j)$ is also an elementary set. Unfortunately, this family is not closed under the complement operation, because we do not allow unbounded intervals, but it is 'almost' closed under the complement operation.

Lemma 1.6. *If R and S are boxes, then $R - S = R \cap S^c$ is a finite union of rectangles.*

Corollary 1.7. *If E and F are elementary sets, then $E - F$ is an elementary set.*

Proof.

$$(R_1 \cup \cdots \cup R_n) \cap (S_1 \cup \cdots \cup S_m)^c = \cup (R_i \cap S_j)$$

$\square$

A family of sets is an **algebra** if it is closed under the union and complement operation. It is easy to see that an algebra is closed under intersections because $E \cap F = (E^c \cup F^c)^c$, as well as the set subtraction operation. What we have just proven is that the set of elementary subsets contained in $[0,1]^n$ is an algebra. The fact that elementary sets are not an algebra is one of the reasons we have to enlarge the family of sets we measure the size of.

Let's look at some elementary properties of the behaviour of μ on elementary sets:

- It is also easy to see that if $E_1, \dots, E_n$ are disjoint elementary sets, then $\mu(E_1 \cup \cdots \cup E_n) = \sum \mu(E_i)$, so μ is *finitely additive* on elementary subsets.

- For any elementary set E, $\mu(E+x) = \mu(E)$, so μ is *translation invariant*.

- It follows that if E and F are elementary sets, with $E \subset F$, then $\mu(E) \leqslant \mu(F)$, because

  $$\mu(F) = \mu((F-E) \cup E) = \mu(F-E) + \mu(E) \geqslant \mu(E)$$

  thus μ is a *monotone* function on sets.

These properties uniquely define the function μ we constructed up to a scalar factor. Since all the properties are intuitive to us, this tells us we're going in the right direction!

1.2 Jordan Measurable Sets

In the last section, we constructed a consistant measure μ on the family of elementary sets. However, this family is certainly limited. We cannot even use this quantity to measure the area of a circle, or the volume of a sphere. However, we have really only applied the first ancient technique of measuring area, forming disjoint unions of simple shapes. We haven't used the method of approximating shapes by simple sets from above and below. If E is an arbitrary set, and there is a constant $C \geqslant 0$ such that for all ε, there are $E_1 \subset E$ and $E \subset E_2$ with

$$C - \varepsilon \leqslant \mu(E_1) \leqslant \mu(E_2) \leqslant C + \varepsilon$$

then it would be reasonable to define the area of E to be C. We call such a set **Jordan measurable**. More specifically, for any subset of $\mathbf{R}^n$, we define

$$\mu_*(E) = \sup\{\mu(E_1) : E_1 \text{ elementary}, E_1 \subset E\}$$

$$\mu^*(E) = \inf\{\mu(E_1) : E_1 \text{ elementary}, E_2 \supset E\}$$

the **inner** and **outer** measures of the set E. We say a *bounded* set E is Jordan measurable if $\mu_*(E) = \mu^*(E)$.

Theorem 1.8. *A set is Jordan measurable if and only if for every $\varepsilon > 0$, there is an elementary set A with $\mu^*(A \triangle E) \leqslant \varepsilon$.*

Proof. Consider a set E satisfying the second condition. Then there is an elementary set F with $A \triangle E \subset F$ and $\mu(F) \leqslant 2\varepsilon$, and so

$$\mu^*(E) \leqslant \mu(A \cup F) \leqslant \mu(A) + \mu(F) \leqslant \mu(A) + 2\varepsilon$$

since $\mu_*(A \triangle E) \leqslant \mu^*(A \triangle E) \leqslant \varepsilon$, we can find $F \subset A \triangle E$ with $\mu(F) \leqslant \varepsilon$, and

$$\mu_*(E) \geqslant \mu_*(A - F) \geqslant \mu_*(A) - \mu_*(F) \geqslant \mu_*(A) - \varepsilon$$

Then we let $\varepsilon \to 0$. Conversely, if E is Jordan measurable, there are elementary sets F_1, F_2 with $F_1 \subset E \subset F_2$ and $\mu(F_2) - \mu(F_1) = \mu(F_2 - F_1) < \varepsilon$. Then

$$\mu^*(F_2 \triangle E) = \mu^*(F_2 - E) \leqslant \mu^*(F_2 - F_1) \leqslant \varepsilon$$

This shows the condition holds, and also that we can choose the set A to be a superset of E in the proof above. ∎

Jordan measurable sets satisfy the same algebraic operations are the family of elementary sets.

- If $\mu^*(A - E), \mu^*(B - F) < \varepsilon$, then

  $$\mu^*((A \cap B) - (E \cap F)) \leqslant \mu^*(A - E) + \mu^*(B - F) \leqslant 2\varepsilon$$

  hence $E \cap F$ is measurable if E and F are measurable.

- If $\mu^*(A-E), \mu^*(B-F) < \varepsilon$, then since

$$(A \triangle B) \triangle (E \triangle F) = (A \triangle E) \triangle (B \triangle F) \subset (A \triangle E) \cup (B \triangle F)$$

  Hence $\mu((A \triangle B) \triangle (E \triangle F)) \leqslant \varepsilon$, and the symmetric difference of two sets is measurable.

- Since $E - F = (E \triangle F) \cap E$, the set theoretic minus of two Jordan measurable sets is Jordan measurable.

- To prove $E \cup F$ is Jordan measurable, we can assume E and F are disjoint, because $(E \cup F) = (E \cap F) \cup (E - F) \cup (F - E)$. Note that in this case $\mu_*(E) + \mu_*(F) \leqslant \mu_*(E + F)$, because an interior estimate of E and an interior estimate of F combine as disjoint sets to give an interior estimate of $\mu_*(E+F)$. Now since E and F are measurable, for any $\varepsilon > 0$, we can find elementary sets $E \subset E^*$ and $F \subset F^*$ such that $\mu(E^*) \leqslant \mu_*(E) + \varepsilon$ and $\mu(F^*) \leqslant \mu_*(F) + \varepsilon$. But then

$$\mu(E^* \cup F^*) \leqslant \mu(E^*) + \mu(F^*) \leqslant \mu_*(E) + \mu_*(F) + 2\varepsilon \leqslant \mu_*(E \cup F) + 2\varepsilon$$

  This shows $E \cup F$ is Jordan measurable.

hence, restricted to a particular bounded Jordan measurable set of space, the class of Jordan measurable sets is an algebra. Since $\mu^*(E \cup F) \leqslant \mu^*(E) + \mu^*(F)$ and $\mu_*(E \cup F) \geqslant \mu_*(E) + \mu_*(F)$ holds for all disjoint sets E and F, we conclude that Jordan measurable sets have finite additivity. This implies monotonicity. The translation invariance of the measure follows from the translation invariance of the measure on elementary sets.

Example. *Let R be a closed box in $\mathbf{R}^n$, and $f : R \to \mathbf{R}$ a continuous function. Then the graph $\Gamma(f)$ of f in $\mathbf{R}^{n+1}$ is Jordan measurable, and has Jordan measure zero. This follows because, since R is compact, f is uniformly continuous, and therefore for any ε there are finitely many disjoint boxes $S_1, \dots, S_m$ covering R such that if two points x and y lie in the same box, $|f(x) - f(y)| < \varepsilon$. But this implies that if we fix a point x_i in S_1, then the sets $S_i \times [x_i - \varepsilon, x_i + \varepsilon]$ cover the graph, and so*

$$\mu^*(\Gamma(f)) \leqslant \sum \mu(S_i \times [x_i - \varepsilon, x_i + \varepsilon]) \leqslant 2\varepsilon \sum \mu(S_i) \leqslant 2\varepsilon\mu(R)$$

let $\varepsilon \to 0$ to obtain that $\Gamma(f)$ has upper measure zero, and thus lower measure zero. The set $X = \{(x,t) : 0 \leqslant f(x) \leqslant t\}$ is also Jordan measurable. Given the *same S_i, the sets $S_i \times [0, x_i - \varepsilon)$ are contained in X, and $S_i \times [0, x_i + \varepsilon]$ contain X, and the difference between these two sets is exactly the sets we used to show the measure of the boundary is zero, hence X is Jordan measurable because it can be approximated from above and below.*

Example. *A triangle is not an elementary set, but it is a Jordan measurable set. First, consider a right triangle with sides parallel to the x and y axis. Then, by a translation, we can write one point as $(x, 0)$ and another as $(0, y)$. For each N, consider the disjoint sequence of rectangles*

$$[0, x/N) \times \{0\} \cup [x/N, 2x/N) \times [0, y/N] \cup \dots \cup [\frac{N-1}{N}x, x] \times [0, \frac{N-1}{N}y]$$

These rectangles are contained in the triangle, and so

$$\mu_*(T) \geqslant \sum_{i=0}^{N-1} \frac{x}{N}\frac{iy}{N} = \frac{xy}{N^2}\frac{(N-1)N}{2} = \frac{N-1}{N}\frac{xy}{2}$$

Letting $N \to \infty$, we find $\mu_*(T) \geqslant xy/2$. On the other hand, consider the disjoint sequence of rectangles

$$[0, x/N) \times [0, y/N] \cup [x/N, 2x/N) \times [0, 2y/N] \cup \dots \cup [\frac{N-1}{N}x, x] \times [0, y]$$

which contains the triangle, so

$$\mu^*(T) \leqslant \sum_{i=1}^{N} \frac{x}{N}\frac{iy}{N} = \frac{N+1}{N}\frac{xy}{2}$$

Letting $N \to \infty$, we conclude $\mu^*(T) \geqslant xy/2$. Equating estimates, we determine the triangle is Jordan measurable with area $xy/2$. If only one of the sides is horizontal, we may split the triangle into two right triangles with the other side perpendicular to the y axis, so this shape is measurable. If one coordinate is $(x, 0)$, and the other

DO MORE EXERCISES FROM TAO'S BOOK

The Jordan measure is intrinsically connected with the Riemann integral. Given an interval,

TALK ABOUT RIEMANN INTEGRAL, SHOW SET IS JORDAN MEASURABLE IFF ITS BOUNDARY HAS MEASURE ZERO?

1.3 Lebesgue Measure

If we are able to stick with Jordan measurable sets, we should, because it is here that integration theory works in the best way. However, not all sets are Jordan measurable, and often when studying fractal sets one runs into unmeasurable sets. What's more, even if a set $E_1, E_2, \dots$ is Jordan measurable, their union $\bigcup E_i$ and their intersection $\bigcap E_i$ need not be measurable, even if these sets are bounded. In terms of Riemann integrability, this causes problems with understanding the pointwise limit of functions: A sequence of uniformly bounded Riemann integrable functions $f_n : [0,1] \to \mathbf{R}$ which converges pointwise to a bounded function $f : [0,1] \to \mathbf{R}$ need not be Riemann integrable. If we replace pointwise convergence with uniform convergence, then f will be Riemann integrable, but this relates to the fact that uniform convergence allows one to conver f with finitely many rectangles (ELABORATE HERE).

To obtain a family of sets with a well defined measure theory which satisfies countable additivity, we must tinker with how we defined Jordan measure. Recall that to obtain Jordan measure, we took outer and inner estimates of arbitrary sets by covers of *finitely many* rectangles. We obtain Lebesgue integrability if we replace the finiteness with infinitely many rectangles. We consider the values

$$\mu^*(E) = \inf\{\textstyle\sum\}$$

Definition. If A is a set of real numbers, then it's Lebesgue measure is

$$m(A) = \inf\left\{ \sum_{k=1}^{\infty} m(I_k) : \bigcup_{k=1}^{\infty} I_k \supset A \right\}$$

The end goal of this passage is to find out what it takes to prove that if $\{A_i\}$ is a disjoint collection of sets, then $m(\bigcup A_i) = \sum m(A_i)$. This will give us intuition in the abstract case. In this case, one side of the equality is fairly easy to show.

Theorem 1.9. *If $\{A_i\}$ is a countable collection of sets, then $m(\bigcup A_i) \leqslant \sum m(A_i)$.*

Proof. If any A_i has infinite length, then the theorem is trivial. Thus assume all A_i have finite measure. Fix some $\varepsilon > 0$. For each A_i, pick a countable set $\mathcal{I}_i$ of open intervals such that $\sum_{I \in \mathcal{I}_i} I \leqslant m(A_i) + \varepsilon/2^k$. Then $\bigcup \mathcal{I}_i$ is a countable collection of open intervals covering $\bigcup A_i$, and so

$$m(\bigcup A_i) \leqslant \sum_{i=1}^{\infty} \sum_{I \in \mathcal{I}} m(I) \leqslant \sum_{i=1}^{\infty} [m(A_i) + \varepsilon/2^k] = \sum_{i=1}^{\infty} m(A_i) + \varepsilon$$

The proof is completed since ε was arbitrary. □

Lemma 1.10. *If $A \subset B$, $m(A) \leqslant m(B)$.*

Proof. Any cover of B is a cover of A. □

Let us check the m is well defined, when passing from lengths of intervals to approximations of arbitrary sets.

Lemma 1.11. *For any interval $I = (a,b)$, $m(I) = b - a$*

Proof. First, we will verify that $m([a,b]) = b - a$. Let $\mathcal{I}$ be a collection of open intervals such that $\bigcup \mathcal{I} \supset [a,b]$. Without loss of generality, we may choose a finite subcover, since $[a,b]$ is compact. Using this finiteness, construct a sequence $(a_1, b_1), \dots, (a_n, b_n)$ from $\mathcal{I}$ such that $b_i \geqslant a_{i+1}$ for each i, $a_1 \leqslant a$, and $b_n \geqslant b$. Then

$$\begin{aligned}
\sum_{I \in \mathcal{I}} m(I) \geqslant \sum_{i=1}^{n} m((a_i, b_i)) &= \sum_{i=1}^{n} b_i - a_i \\
&\geqslant (b_n - a_n) + \sum_{i=1}^{n-1} (a_{i+1} - a_i) \\
&= (b_n - a_n) + (a_n - a_1) = b_n - a_1 \geqslant b - a
\end{aligned}$$

Thus $m([a,b]) \geqslant b - a$. Now, fix $\varepsilon > 0$. Choose the cover

$$\mathcal{I} = \{(a - \varepsilon, a + \varepsilon), (a,b), (b - \varepsilon, b + \varepsilon)\}$$

Now $\bigcup \mathcal{I} = (a - \varepsilon, b + \varepsilon) \supset [a,b]$, so

$$m([a,b]) \leqslant m((a,b)) + m((a - \varepsilon, a + \varepsilon)) + m((b - \varepsilon, b + \varepsilon)) = b - a + 4\varepsilon$$

Since ε was arbitrary, $m([a,b]) \leqslant b - a$, and so $m([a,b]) = b - a$.

Surely, $m((a,b)) \leqslant m([a,b]) = b-a$. But also, by Lemma (1.1),

$$m([a,b]) \leqslant m((a,b)) + m(\{a\}) + m(\{b\}) = m((a,b))$$

since the length of a single point is zero. $\square$

Now we want to know that measuring the union is the same as measuring the component parts, as our intuition would tell us. However, Banach and Tarski have warned us that this won't be true of all sets. One side of the inequality can be shown for all sets, but we must specialize to obtain equality – defining exactly what it means for a set to be measurable, as we were discussing above.

Definition. A set A is **measurable** (in the manner of Lebesgue), if for any other set B, $m(B) = m(A \cap B) + m(A^c \cap B)$.

It is simple to verify that $\mathbf{R}$ is a measurable set, and if A is measurable, then so is A^c. More complicated is the fact that open intervals are measurable.

Lemma 1.12. *For any real number a, (a,∞) is measurable.*

Proof. Let A be an arbitrary set. Let $\mathcal{I}$ be a countable collection of intervals such that $\sum_{I \in \mathcal{I}} m(I) \leqslant m(A) + \varepsilon$. Then, for each I, either $I \cap (a,\infty)$ is empty or an interval, as is $I \cap (-\infty, a]$, and the measure of I is equal to the measure of the sum. Thus

$$m(A \cap (a,\infty)) + m(A \cap (-\infty,a]) \leqslant \sum m(I_k \cap (a,\infty)) + \sum m(I_k \cap (-\infty,a]) = \sum m(I_k) \leqslant m(A) + \varepsilon$$

So $m(A \cap (a,\infty)) + m(A \cap (-\infty,a]) \leqslant m(A)$, and we have already proved the inequality the other way. $\square$

Lemma 1.13. *If A and B are measurable, then so is $A \cup B$.*

Proof. Let S be an arbitrary subset of the reals. Then

$$\begin{aligned}
m(S) &= m(S \cap A) + m(S \cap A^c) \\
&= m(S \cap A) + m(S \cap A^c \cap B) + m(S \cap A^c \cap B^c) \\
&= m(S \cap A) + m(S \cap B \cap A^c) + m(S \cap [A \cup B]^c) \\
&= m([S \cap A] \cup [S \cap B \cap A^c]) + m(S \cap [A \cup B]^c) \\
&= m(S \cap [A \cup B]) + m(S \cap (A \cup B)^c)
\end{aligned}$$

One may get from the second last equation to the third last equation by applying the measurability of A to the first measured set. □

Corollary 1.14. *If A and B are measurable, then $A \cap B$ and $A - B$ are measurable.*

Proof. $A \cap B = (A^c \cup B^c)^c$, and $A - B = A \cap B^c$. □

Corollary 1.15. *All open intervals are measurable.*

Corollary 1.16. *If A is any set, and $E_1, \dots, E_n$ is a finite collection of disjoint measurable sets, then*

$$m(A \cap (\bigcup_{k=1}^{n} E_k)) = \sum_{k=1}^{n} m(A \cap E_k)$$

What we have shown here is that the set of measurable sets is a Boolean algebra. We can go one further.

Lemma 1.17. *If $E_1, E_2, \dots$ is a countable collection of disjoint measurable sets, then $E = E_1 \cup E_2 \cup \dots$ is measurable.*

Proof. Define $F_n = \bigcup_{k=1}^{n} E_n$. Then F_n is measurable, and $F_n^c \supset E^c$. Hence

$$\begin{aligned} m(A) &= m(A \cap F_n) + m(A \cap F_n^c) \geqslant m(A \cap F_n) + m(A \cap E^c) \\ &= \sum_{k=1}^{n} m(A \cap E_k) + m(A \cap E^c) \end{aligned}$$

Since n was arbitrary,

$$m(A) \geqslant \sum_{k=1}^{\infty} m(A \cap E_k) + m(A \cap E^c) \geqslant m(A \cap E) + m(A \cap E^c)$$

But $m(A) \leqslant m(A \cap E) + m(A \cap E^k)$, so E is measurable. □

Corollary 1.18. *The countable union of measurable sets are measurable.*

Proof. Simply modify measurable sets by elementary set operations so they are disjoint, and then take their union. □

From this theorem, we can determine that every open set in **R** is open, as every open set is the countable union of open intervals.

We can know show what we originally set out to solve.

Theorem 1.19. *If $\{A_i\}$ is a countable collection of pairwise disjoint measurable sets, then*

$$m(\bigcup A_i) = \sum m(A_i)$$

Proof. The calculations in the above theorem show that, letting $A = \bigcup A_i$,

$$m(A) \geqslant \sum_{k=1}^{\infty} m(A \cap A_i) + m(A \cap A^c) = \sum_{k=1}^{\infty} m(A_i)$$

but this shows equality, since the other direction of inequality always holds. □

The function m, restricted to measurable sets, will now be known as the Lebesgue measure on **R**. It is the first in a line of a general class of functions known as measures, defined on subsets of a space and measuring these subset's size. The notions of measurable set will be abstracted to the properties proved above. That is, we can measure the union, intersection, and complement of all measurable sets. Most theorems in measure theory are actually be proved in general just as easily as on the Lebesgue measure we have just described.

1.4 Appendix: Banach Tarski

Let us consider the sphere. A nice property of this object is that it is invariant under any rotation - that is, if you take a point, and rotate it around the origin, you will never end up at a point off the unit sphere. Mathematically, we say that the orthogonal group $O(3)$ acts on the sphere S^1.

The core technique of this proof can be executed in a simpler form on free groups. Consider the free group $F_{\{a,b\}}$ on two characters. Let $S(a)$ be the set of all sequences whose simplest form begins with a, and define $S(b)$, $S(a^{-1})$, and $S(b^{-1})$ similarily. We have the following equalities:

$$F_{\{a,b\}} = S(a) \cup S(b) \cup S(a^{-1}) \cup S(b^{-1})$$

$$F_{\{a,b\}} = S(a) \cup aS(a^{-1})$$

$$F_{\{a,b\}} = S(b) \cup bS(b^{-1})$$

Thus we have partitions $F_{\{a,b\}}$ into four sets. By 'rotating' two of these partitions, we obtain two copies of the group.

The trick to the Banach-Tarski paradox on the sphere is to find subsets of the orthogonal group that behave like $F_{\{a,b\}}$. We will say a subset X of euclidean space can be **paradoxically decomposed**, if it can be expressed as the disjoint union of subsets, $X_1 \cup \cdots \cup X_n$, and, under group actions $g_1, \ldots, g_n \in O(3)$, we may express $X = g_1 X_1 \cup \cdots \cup g_i X_i$, and $X = g_{i+1} X_{i+1} \cup \cdots \cup g_n X_n$, for some i.

Lemma 1.20. *There is a subgroup of $O(3)$ isomorphic to $F_{\{a,b\}}$.*

Proof. Map a to a rotation horizontally by $\sqrt{2}\pi$ radians, and map b to a rotation vertically by $\sqrt{2}\pi$ radians. This induces a homomorphism from $F_{\{a,b\}}$ to $O(3)$. We claim this homomorphism is injective. □

Theorem 1.21 (Banach Tarski)**.** *The sphere may be paradoxically decomposed.*

Chapter 2

Abstract Measures

Let us now begin to describe measures in their abstract generality. Let X be an arbitrary set. A σ-algebra on X is a family of subsets of X, called measurable sets, such that X is measurable, if A is measurable, then A^c is measurable, and the union of any *countable* family of sets is measurable.

Chapter 3

Measure Extension

To construct the Lebesgue measure, we began with an intuitive notion of area over the set of intervals on the real line, and then extended the notion of length to a bigger class of sets, the measurable sets. This chapter attempts to purify this strategy to work on arbitrary measure spaces. The idea is to take a countably additive measure defined on a family of sets satisfying weaker conditions than that of a σ algebra, and then to extend this measure to the σ algebra this family generates.

If X is a set, a family Σ_0 of subsets of X is known as an **algebra** if it is closed under complements and finite unions (it is an algebra under $\mathbf{F}_2$, if we consider intersection as multiplication, and the *symmetric difference* $A \triangle B = (A-B) \cup (B-A)$ as addition). A σ algebra is just an algebra where we can take countable unions instead of finite unions, and this is the natural place to study the analytical theory of measure, because we often want to apply limiting estimates.

The way we constructed the Lebesgue measure was to take a family of elementary sets (disjoint unions of boxes), which we intuitively know the measure of, and then define the measure of arbitrary sets by approximating them from the outside by these elementary sets. The family of elementary sets formed an algebra, but not a σ algebra, and thus the extension was necessary to construct the satisfying theory of the Lebesgue measure. To generalize this construction, we take a algebra Σ_0, on which a sense of measure is defined, and try to extend this to a σ algebra containing Σ_0. The intuitive notion of measure is an additive function $\mu_0 : \Sigma_0 \to [0, \infty]$ such that $\mu_0(\varnothing) = 0$, which is countably additive when countable additivity is well defined in Σ_0. We call such functions **pre-measurable**. Alternatively, a premeasure is a monotone, countably subadditive function where defined, or a monotone, upward continuous function. Being a premeasure is the weakest condition we could put on a function which is a candidate for extension to a full measure, but Caratheodory proved that the condition is sufficient for extension.

To obtain an extension, we employ the same ideas used in the construction of the Lebesgue measure. In that situation, we started with a simple algebra of sets (unions of boxes), and constructed the upper approximations

$$\mu^*(S) = \inf \left\{ \sum_{k=0}^{\infty} \mu(A_k) : A_k \in \Sigma_0, S \subset \bigcup A_k \right\}$$

which are defined for any subset of $\mathbf{R}^d$. This function is monotone, countably σ subadditive, and $\mu^*(\emptyset) = 0$. On any set X, we call a function μ^* on X of this form an **exterior measure**. One of Caratheodory's major contribution to the theory of measure is to notice that a set E is Lebesgue measurable if for any subset A of $\mathbf{R}^d$,

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

For a general exterior measure, we say a set E is **Caratheadory measurable** if it satisfies every equation of the form above. Since μ^* is countably subadditive, it suffices to verify that $\mu^*(A)$ upper bounds the sum of the other two sets. This implies every set of exterior measure zero is Caratheadory measurable.

Lemma 3.1. *The set of Caratheadory measurable sets forms a σ algebra, and μ^* is a measure on this family.*

Proof. Clearly $\emptyset$ and X are Caratheadory measurable, and the symmetry of the statement implies the class of Caratheadory measurable sets is closed under complements. Next, we show measurable sets are closed under finite unions of disjoint sets, and that μ^* is finitely additive. If E_1 and E_2 are disjoint and measurable, and A is arbitrary, then

$$\begin{aligned} \mu^*(A) &= \mu^*(E_1 \cap A) + \mu^*(E_1^c \cap A) \\ &= \mu^*(E_1 \cap E_2 \cap A) + \mu^*(E_1 \cap E_2^c \cap A) \\ &+ \mu^*(E_1^c \cap E_2 \cap A) + \mu^*(E_1^c \cap E_2^c \cap A) \\ &\geqslant \mu^*((E_1 \cup E_2) \cap A) + \mu^*((E_1 \cup E_2)^c \cap A) \end{aligned}$$

This shows measurability, and

$$\mu^*(E_1 \cup E_2) = \mu^*((E_1 \cup E_2) \cap E_2) + \mu^*((E_1 \cup E_2) \cap E_2^c) = \mu^*(E_2) + \mu^*(E_1)$$

Finally, we need to show that the class of Caratheadory measurable sets is closed under countable unions of disjoint sets, and μ^* is countably additive on this family. Consider measurable sets $E_1, E_2, \dots$, and let

$$K_n = \bigcup_{k \leqslant n} E_k \quad K = \bigcup E_k$$

Each K_n is measurable, and $\mu^*(K_n) = \sum_{k \leqslant n} \mu^*(E_k)$. If A is arbitrary, by induction, we find

$$\mu^*(K_n \cap A) = \sum_{k=1}^{n} \mu^*(E_n + A)$$

and thus

$$\mu^*(A) = \mu^*(K_n \cap A) + \mu^*(K_n^c \cap A) \geqslant \sum_{k=1}^{n} \mu^*(E_n \cap A) + \mu^*(K^c \cap A)$$

Letting $n \to \infty$ gives that

$$\mu^*(A) \geqslant \sum_{n=1}^{\infty} \mu^*(E_n \cap A) + \mu^*(K^c \cap A) \geqslant \mu^*(K \cap A) + \mu^*(K^c \cap A)$$

This shows K is measurable, and taking $A = K$ in the calculation with an equality gives countable additivity. □

Caratheodory's theorem generates measure spaces that are **complete**, in the sense that any set contained within a set of measure zero also has measure zero. The lemma leads to a quick proof of the extension theorem.

Theorem 3.2 (Caratheodory's Extension Theorem)**.** *Every premeasure on an algebra Σ_0 can be extended uniquely to a measure on the σ algebra Σ that Σ_0 generates.*

Proof. Given μ_0, define the exterior measure

$$\mu^*(S) = \inf \left\{ \sum_{k=1}^{\infty} \mu(A_k) : A_k \in \Sigma_0, S \subset \bigcup A_k \right\}$$

It is easy to verify this is an exterior measure. Monotonicity is easy, and to prove countable subadditivity, note that if $S_1, S_2, \dots$ are a family of sets with union S and $\sum \mu^*(S_i) < \infty$, then we can find sets $A_1^i, A_2^i, \dots, A_{k_i}^i$ covering S_i with

$$\mu^*(S_i) + \varepsilon/2^n \geqslant \sum \mu(A_k^i)$$

Then the union over all A_j^i is a cover of S, and then

$$\mu^*(S) \leqslant \sum \mu(A_j^i) = \sum \mu^*(S_i) + \varepsilon/2^n \leqslant \sum \mu^*(S_i) + \varepsilon$$

We can then let $\varepsilon \to 0$. For any $S \in \Sigma_0$, $\mu^*(S) = \mu * (S)$, because if $S \subset \bigcup A_k$, where $A_k \in \Sigma_0$, then by monotonicity, $\mu(A_k \cap S) \leqslant \mu(A_k)$, and because

$$S = \lim_{n \to \infty} S \cap \bigcup_{k \leqslant n} A_k$$

we conclude

$$\mu(S) = \lim_{n \to \infty} \mu \left(S \cap \bigcup_{k \leqslant n} A_k \right) \leqslant \lim_{n \to \infty} \mu \left(\bigcup_{k \leqslant n} A_k \right) \leqslant \sum_{k=0}^{\infty} \mu(A_k)$$

This shows in particular that $\mu^*(\varnothing) = \mu(\varnothing) = 0$, so μ^* is an exterior measure. Because of Caratheadory's lemma, all that remains is to show that all elements E of Σ_0 are Caratheadory measurable. Given A, first assume $\mu^*(A) < \infty$. Then we can find a family of sets $A_1, A_2, \dots \in \Sigma_0$ with $\sum \mu(A_k) \leqslant \mu^*(A) + \varepsilon$. But then

$$\sum \mu(A_k) = \sum \mu(A_k \cap E) + \mu(A_k \cap E^c)$$

The sets $A_k \cap E$ cover $A \cap E$, and the sets $A_k \cap E^c$ cover $A \cap E^c$, so

$$\mu^*(A \cap E) + \mu^*(A \cap E^c) \leqslant \sum \mu(A_k \cap E) + \mu(A_k \cap E^c) = \sum \mu(A_k) \leqslant \mu^*(A) + \varepsilon$$

We then let $\varepsilon \to 0$. If $\mu^*(A) = \infty$, $\mu^*(A) \geqslant \mu^*(A \cap E) + \mu^*(A \cap E^c)$ is obvious. This shows that the σ algebra of Caratheadory measurable sets contains Σ_0, and the measure μ^* defined on it extends μ. Note, however, that there may be many more Caratheadory measurable sets than the subsets of Σ, but we can solve the problem by restriction. $\square$

Corollary 3.3. *If X is σ-finite, the extension is unique on Σ.*

Proof. Let μ be another extension of μ_0 to Σ rather than μ^*. Choose some $E \in \Sigma$, and fix $\varepsilon > 0$. For any countable covering of E by elements $A_1, \dots \in \Sigma_0$,

$$\mu(A) \leqslant \sum \mu(A_k) = \sum \mu_0(A_k)$$

It follows that $\mu \leqslant \mu^*$. By symmetry, $\mu(A^c) \leqslant \mu^*(A^c)$. But then $\mu(A) + \mu(A^c) = \mu(X) = \mu^*(X)$, and assuming $\mu(X) < \infty$, we conclude

$$\mu(A) = \mu(X) - \mu(A^c) \geqslant \mu^*(X) - \mu^*(A^c) = \mu^*(A)$$

In general, if $X = \lim X_n$, where X_n are finite measure sets, we can apply the theorem to $E \cap X_n$, and then take limits. □

For the purposes of using the Caratheodory extension theorem, is is often useful to use the following method of generating a premeasure. A **semialgebra** on a set X is a family subsets containing the empty set, closed under intersections, and such that if A is in the family, then A^c can be broken into finitely many disjoint sets in the family. The algebra generated by a semialgebra is then the family of all disjoint unions of elements of the semialgebra. A content is a $[0, \infty]$ valued function μ defined on a semialgebra with $\mu(\emptyset) = 0$ and $\mu(A \cup B) = \mu(A) + \mu(B)$ for any two disjoint A, B in the semialgebra. Any content can then be extended uniquely to a premeasure on the algebra generated by the semialgebra by taking sums of disjoint unions.

Example. *Let us consider the construction of the Lebesgue measure in the form of the Caratheodory extension theorem. Consider the semialgebra of intervals of the form $[a, b)$ and $(-\infty, b)$, where we allow open brackets to take the value ∞. Define*

$$\mu[a, b) = b - a \quad \mu(-\infty, b) = \infty$$

This extends to the algebra of disjoint unions of intervals. It is nontrivial to verify that μ is countably additive, which we partook in the first chapter, but once this is done, the Carathedory extension theorem provides a black box to extend μ to the Lebesgue measure on $\mathbf{R}$. More generally, we can consider the semialgebra of boxes with open and closed boundaries in $\mathbf{R}^d$, and the resulting measure we construct from this family will be the Lebesgue measure in $\mathbf{R}^d$.

Chapter 4

Product Spaces

Let X and Y be measure spaces, with respective σ algebras Σ and Π. The idea of product spaces is to define a natural measure space structure on $X \times Y$. We define an **elementary rectangle** in $X \times Y$ to be a set of the form $E \times F$, where $E \in \Sigma$ and $F \in \Pi$. We shall find the natural σ algebra on $X \times Y$ is the one generated by elementary rectangles, and we denote this σ algebra by $\Sigma \otimes \Pi$ (it really is the tensor product viewing the two σ algebras as actual algebras). Since

$$(E_1 \times F_1) \cap (E_2 \cap F_2) = (E_1 \cap E_2) \times (F_1 \cap F_2)$$

we conclude the family of elementary rectangles is a π system. Given a set $E \subset X \times Y$, we define the *sections*

$$E_x = \{y : (x,y) \in E\} \qquad E^y = \{x : (x,y) \in E\}$$

These sections preserve measurability.

Lemma 4.1. *If E is a measurable subset of $X \times Y$, then E_x and E^y are measurable subsets of X and Y.*

Proof. Fix $y \in Y$. If $E \times F$ is an elementary rectangle, then if $y \in F$, $(E \times F)^y = E$, and if $y \notin F$, $(E \times F)^y = \emptyset$, and these sets are both measurable in X. If E^y and F^y are measurable subsets of Y, then $(F - E)^y = F^y - E^y$ is measurable. If $E = \bigcup E_i$ is the countable union of sets with E_i^y measurable, then $E^y = \bigcup E_i^y$ is also measurable. We conclude the class of sets E with E^y measurable is a λ system containing the π system of elementary rectangles, so the π-λ theorem guarantees the theorem is true for all measurable subsets of $X \times Y$. The symmetry of the situation shows that E_x is measurable for all measurable sets. $\square$

It follows that if $f : X \times Y \to Z$ is a measurable function, then the functions $f^y : x \mapsto f(x,y)$ and $f_x : y \mapsto f(x,y)$ are also measurable, because they are the composition of f with the maps $x \mapsto (x,y)$, $y \mapsto (x,y)$, and these project backward onto the sections of measurable sets in $X \times Y$.

The $\pi - \lambda$ theorem isn't often used in measure theory, where it is often substituted for the monotone class theorem. We say a family of sets Σ is a **monotone class** if it is closed under upward limits: If $E_1 \subset E_2 \subset \dots \in \Sigma$, then $\lim E_i \in \Sigma$, and if $E_1 \supset E_2 \supset \dots \in \Sigma$, then $\lim E_i \in \Sigma$.

Theorem 4.2. *The smallest monotone class containing an algebra is the σ algebra generated by the algebra.*

Proof. If Σ is an algebra, then it is a π system. The $\pi - \lambda$ theorem says that the smallest λ system containing Σ is the σ algebra containing the π system. But if Σ_* is a monotone class containing Σ, then it is almost a λ system, except that we may not have $B - A \in \Sigma_*$ if $A \subset B$ are both members of Σ_*. We show that this does hold if Σ_* contains Σ. For any set E, let

$$X_E = \{F : E - (E \cap F) \in \Sigma_*\} \qquad Y_E = \{F : F - (E \cap F) \in \Sigma_*\}$$

For any set E, X_E and Y_E are monotone sets, which essentially follows because if $F_i \to F$ (upward or downward), then $E - (E \cap F_i) \to E - (E \cap F)$ and $F_i - (E \cap F_i) \to F - (E \cap F)$. Now if $E \in \Sigma$, then X_E and Y_E both contain Σ, so $\Sigma_* \subset X_E, Y_E$. But this means if $E \in \Sigma_*$, $F \in \Sigma$, then $E \in X_F$, so $F - (E \cap F) \in \Sigma_*$. This implies that $F \in Y_E$. We conclude if $E \in \Sigma_*$, Y_E is a monotone set containing Σ, so $\Sigma_* \subset Y_E$. This shows that Σ_* is a λ system, and therefore that Σ_* contains the σ algebra generated by Σ. $\square$

We can use the monotone class theorem to understand the structure of the product σ algebra $\Sigma \times \Pi$ in more detail. We know that the set of all elementary rectangles $E \times F$ is closed under intersection, but isn't closed under the other algebraic operations which would allow us to use the monotone class theorem. Nonetheless, if we consider the class of all **elementary sets**, which are subsets of $X \times Y$ formed from *disjoint unions* of elementary rectangles, then this does form an algebra. The essential reason for this is that

$$(E_1 \times F_1) - (E_2 \times F_2) = (E_1 - E_2) \times F_1 \cup (E_1 \cap E_2) \times (F_1 - F_2)$$

$$(E \times F)^c = E^c \times F \cup E \times F^c \cup E^c \times F^c$$

$$(E_1 \times F_1) \cup (E_2 \times F_2) = (E_1 \times F_1) \cup [(E_2 \times F_2) - (E_1 \times F_1)]$$

and the fact that algebraic operations distribute themselves over unions. Since the family of elementary sets *does* form an algebra, we conclude that $\Sigma \otimes \Pi$ is the smallest monotone class which contains the family of elementary sets. It is of course also useful to note that $\Sigma \otimes \Pi$ is the smalles the λ system containing the π system of elementary rectangles.

Now suppose we have a positive measure μ defined on Σ, and a positive measure λ on Π. It makes sense to define a measure $\mu \otimes \lambda$ on $\Sigma \otimes \Pi$, such that $(\mu \otimes \lambda)(E \times F) = \mu(E)\lambda(F)$, just like we would define the area of a rectangle based on the length of its sides. If $E_1 \times F_1$ and $E_2 \times F_2$ are two disjoint elementary rectangles whose union is also an elementary rectangle $E_3 \times F_3$, then one finds that either E_1 is disjoint from E_2, and their union is E_3, and $F_1 = F_2 = F_3$, or $E_1 = E_2 = E_3$ and F_1 is disjoint from F_2. In the first case, one find

$$(\mu \otimes \lambda)(E_1 \times F_1) + (\mu \otimes \lambda)(E_2 \times F_2) = [\mu(E_1) + \mu(E_2)]\lambda(F_3) = \mu(E_3)\lambda(F_3)$$

and in the second

$$(\mu \otimes \lambda)(E_1 \times F_1) + (\mu \otimes \lambda)(E_2 \times F_2) = \mu(E_3)[\lambda(F_1) + \lambda(F_2)] = \mu(E_3)\lambda(F_3)$$

so $(\lambda \otimes \mu)$ is a content, and thus extends to an additive measure on the family of elementary sets in $\Sigma \otimes \Pi$. Now if a rectangle $E \times F$ is a disjoint, countable union of disjoint rectangles $E_i \times F_i$, we claim that

$$(\mu \otimes \lambda)(E \times F) = \sum (\mu \otimes \lambda)(E_i \times F_i) = \sum \mu(E_i)\lambda(F_i)$$

Fixing $x \in E$, for each $y \in F$ there is exactly one index i with $x \in E_i$ and $y \in F_i$.The countable additivity of λ therefore guarantee that

$$\chi_E(x)\lambda(F) = \sum_{j=1}^{\infty} \chi_{E_j}(x)\lambda(F_j)$$

Applying the monotone convergence theorem over Y, we conclude that

$$\mu(E)\lambda(F) = \int \chi_E(x)\lambda(F)d\mu(x) = \sum_{j=1}^{\infty} \int \chi_{E_j}(x)\lambda(F_j) = \sum_{j=1}^{\infty} \mu(E_j)\lambda(F_j)$$

and this is exactly countable additivity. This shows that $(\lambda \otimes \mu)$ is a *pre-measure* on the class of elementary sets, and the Caratheadory extension theorem guarantees that $\lambda \otimes \mu$ extends to a measure on $\Sigma \times \Pi$. Provided that we are working over a σ finite space, this extension is unique, a fact that will become increasingly more important when we analyze the integration theory of product spaces, where we often have to assume we are working over σ finite spaces.

Theorem 4.3. *If X and Y are σ finite spaces, and $E \in \Sigma \otimes \Pi$, then*

$$\int \lambda(E_x) d\mu(x) = (\lambda \otimes \mu)(E) = \int \mu(E^y) d\lambda(y)$$

Proof. The theorem is trivially true if E is an elementary rectangle. The class of sets for which this theorem holds is also closed monotonely upward, because if $E_1 \subset E_2 \subset \cdots \to E$, then $\lambda((E_i)_x) \to \lambda(E_x)$ monotonely upward, and $\mu(E_i^y) \to \mu(E^y)$ monotonely upward, so the monotone convergence theorem guarantees that if the theorem holds for the E_i, it also holds for E. The hard part of the theorem is showing that if $E_1 \supset E_2 \supset \cdots \to E$, and E_i satisfies the properties of the theorem, then E satisfies the properties of the theorem. Assume first that X and Y are finite measure spaces. Then we can apply the dominated convergence theorem downward, because constant functions are in L^1, and $\lambda(E_x) \leqslant \lambda(Y)$, $\mu(E^y) \leqslant \mu(X)$, and this shows that the class of sets for which the theorem holds is monotone, and contains all elementary rectangles, hence containing all elements of $\Sigma \times \Pi$. To obtain the theorem for σ finite measure spaces, we apply the theorem for any set $E \in \Sigma \times \Pi$ to conclude that if $X_i \subset X$ and $Y_i \subset Y$ have finite measure, then

$$\int_{X_i} \lambda(E_x \times Y_i) = (\lambda \otimes \mu)(E \cap (X_i \times Y_i)) = \int_{Y_i} \mu(E^y \times X_i) d\lambda(y)$$

Now we let $X_i \to X$, $Y_i \to Y$, and apply monotone convergence. $\square$

Theorem 4.4 (Tonelli)**.** *If X and Y are σ finite, and f is a non-negative measurable function, then*

$$\int \int f^y(x) d\mu(x) d\lambda(y) = \int f(x,y) d(\lambda \otimes \mu)(x,y) = \int \int f_x(y) d\lambda(y) d\mu(x)$$

Proof. We have already proven this theorem in the case that f is the indicator function of some measurable subset of $X \times Y$. If f and g satisfy the theorem, then $\alpha f + \beta g$ also satisfy the theorem, for $\alpha, \beta \geqslant 0$. The monotone convergence theorem guarantees the theorem is true for all monotone upward limits of functions for which the theorem is true. But this shows that the theorem is true for all non-negative measurable functions, which are the monotone pointwise limit of simple functions. □

Corollary 4.5. *If f is a complex measurable function on $X \times Y$, where X and Y are both σ finite, then f is in $L^1(X \times Y)$ if and only if the common value of*

$$\int_Y \int_X |f(x,y)| d\mu(x) d\lambda(y) = \int_X \int_Y |f(x,y)| d\lambda(y) d\mu(x)$$

is finite, and this value is equal to $\|f\|_1$.

Theorem 4.6 (Fubini)**.** *If $f \in L^1(X \times Y)$, then $f_x \in L^1(Y)$ for almost all x, $f^y \in L^1(X)$ for almost all y, and*

$$\int \int f_x(y) d\lambda(y) d\mu(x) = \int f(x,y) d(\mu \times \lambda)(x,y) = \int \int f^y(x) d\mu(x) d\lambda(y)$$

Proof. Without loss of generality, assume f is real valued. Then

$$\int_X \int_Y |f(x,y)| d\lambda(y) d\mu(x) < \infty$$

Hence we conclude that for almost all y,

$$\int_Y |f(x,y)| d\lambda(y) = \int_Y |f_x(y)| d\lambda(y) < \infty$$

The same process given in the other direction shows that $|f^y| \in L^1$ for almost all y. To obtain Fubini's integral formula, we just write $f = f^+ - f^-$, and then apply Tonelli's theorem independently to f^+ and f^-. □

4.1 Completion of Product Measures

Just because Σ and Π are complete σ algebras with respect to μ and λ, does not imply that $\Sigma \otimes \Pi$ is complete with respect to $\mu \otimes \lambda$. For instance, in

$\mathbf{R}^2$, if E is a non-measurable subset of $\mathbf{R}$, then $E \times [0,1]$ is not measurable with respect to the twofold product of the Lebesgue measure on $\mathbf{R}$, yet it is measurable with respect to the products completion, because it has exterior measure zero. On the other hand, $E \times [0,1]$ is *Lebesgue* measurable.

Theorem 4.7. *The Lebesgue measure on $\mathbf{R}^{n+m}$ is the completion of the product of the Lebesgue measures on $\mathbf{R}^n$ and $\mathbf{R}^m$.*

Proof. It is easy to see that the Lebesgue measure on $\mathbf{R}^{n+m}$ agrees with the product of the Lebesgue measure $\lambda^n \otimes \lambda^m$ on any set $E \times F$, where E and F are disjoint unions of rectangles in $\mathbf{R}^n$ and $\mathbf{R}^m$. But this is a π system generating the Borel σ algebra of $\mathbf{R}^{n+m}$, so the Lebesgue measure on $\mathbf{R}^{n+m}$ agrees with $\lambda^n \otimes \lambda^m$ on any Borel measurable subset of $\mathbf{R}^{n+m}$. But the Lebesgue measurable subsets of $\mathbf{R}^{n+m}$ are exactly those obtained from the completion of the corresponding measure on the Borel measurable subsets, so this shows every Lebesgue measurable subset is measurable with respect to the completion of the product algebra and vice versa, and the measures agree here. $\square$

Fubini's theorem takes essentially the same form in the complete measure spaces, except for a slight variation.

Theorem 4.8. *If X and Y are complete σ finite measure spaces with σ algebras Σ and Π, and $(\Sigma \otimes \Pi)^*$ is the completion of the product measure on $X \times Y$. If f is $(\Sigma \otimes \Pi)^*$ measurable, then f_x is Σ measurable for almost all x, f^y is Π measurable for almost all y, but the integrals*

$$\int \int f(x,y) d\mu(x) d\lambda(y) \quad \int \int f(x,y) d\lambda(y) d\mu(x)$$

are still well defined, since $\int f(x,y) d\mu(x)$ and $\int f(x,y) d\lambda(y)$ are well defined almost everywhere. Fubini's theorem continues to hold in this circumstance.

Lemma 4.9. *If μ is a positive measure on Σ, and f is Σ^* measurable, then there is a Σ measurable function g with $f = g$ almost everywhere.*

Proof. Suppose $f \geqslant 0$. There there are non-negative Σ^* measurable simple functions $s_0 \leqslant s_1 \leqslant \dots$ converging to f. Then $f = \sum (s_{n+1} - s_n)$. Since $s_{n+1} - s_n$ is a finite combination of characteristic functions, it follows that

$$f(x) = \sum c_i \chi_{E_i}(x)$$

with $c_i > 0$. We know that there are Σ measurable subsets $A_i \subset E_i \subset B_i$ with $\mu(B_i - A_i) = 0$. The function

$$g(x) = \sum c_i \chi_{A_i}(x)$$

is Σ measurable, and $f = g$ except on

$$\bigcup \{E_i - A_i\}$$

which is a countable union of subsets of measure zero. The general case is now immediate. $\square$

Lemma 4.10. *If f is $(\Sigma \times \Pi)^*$ measurable functions on $X \times Y$ such that $f = 0$ almost everywhere with respect to $\mu \otimes \pi$, then for almost all $x \in X$, $f(x,y) = 0$ for almost all y. In particular, f_x is Σ measurable for almost all $x \in X$. Similar results hold for f^y.*

Proof. Let $E = \{(x,y) : f(x,y) \neq 0\}$. Then we know $(\mu \otimes \lambda)(E) = 0$, so there is $F \in \Sigma \otimes \Pi$ with $E \subset F$, and $(\mu \otimes \lambda)(F) = 0$. By Fubini's theorem,

$$\int_X \int_Y \chi_F(x,y) d\lambda(y) d\mu(x) = \int_X \lambda(F_x) d\mu(x) = 0$$

This means that for almost all x, $\lambda(F_x) = 0$, and since $E_x \subset F_x$, this implies $E_x \in \Pi$, because Π is complete, and $\lambda(E_x) = 0$. If $y \notin E_x$, then $f(x,y) = 0$, so $f(x,y) = 0$ for almost all y. $\square$

It is useful sometimes to note that $(\Sigma \otimes \Pi)^* = (\Sigma^* \otimes \Pi^*)^*$. To see this, it suffices to show that if $E \in \Sigma^*$ and $F \in \Pi^*$, then $E \times F \in (\Sigma \otimes \Pi)^*$. We note there are $E_1, E_2 \in \Sigma$, $F_1, F_2 \in \Pi$ with $E_1 \subset E \subset E_2$, $F_1 \subset F \subset F_2$, and $\mu(E_2 - E_1) = \lambda(F_2 - F_1) = 0$, and since

$$\begin{aligned} E_2 \times F_2 - E_1 \times F_1 = (E_2 - E_1) \times (F_2 \cap F_1) \\ \cup (E_2 \cap E_1) \times (F_2 - F_1) \\ \cup (E_2 - E_1) \times (F_2 - F_1) \end{aligned}$$

is the union of three sets with negligable measure in $\mu \otimes \lambda$.

4.2 Integration in Polar Coordinates

It is well known that the space $\mathbf{R}^d - \{0\}$ can be uniquely expanded into polar coordinates in $\mathbf{R}^+ \times S^{d-1}$, by writing $x = r\hat{x}$, where $r = |x|$ and $\hat{x} = x/|x|$. In this section we argue that

$$\int_{\mathbf{R}^d} = \int_{S^{d-1}} \int_0^\infty f(r\hat{x}) r^{d-1} dr d\hat{x}$$

when S^{d-1} has an appropriate measure, and f is not especially eccentric.

The first step is to define the measure space S^{d-1}. We let E be **surface measurable** precisely when it's associated unit sector

$$S_E = \{x \in B^d : \hat{x} \in E\}$$

is Lebesgue measurable. We then define the surface measure σ on the measurable subsets of S^{d-1} by setting $\sigma(E) = d \cdot S(E)$. This is exactly the *pushforward measure* induced from the projection $x \mapsto x/|x|$ from $B^d - \{0\}$ to S^{d-1}, where $B^d - \{0\}$ is given the Lebesgue measure on the Lebesgue measurable subsets. Now define a measure μ on the Lebesgue measurable subsets $(0,\infty)$ by the equation $d\mu = r^{d-1} dr$.

Lemma 4.11. *If $\pi : (0,\infty) \times S^{d-1} \to \mathbf{R}^{d-1}$ is defined by $\pi(r,x) = rx$, then π is Borel measurable, and $\pi_*(\mu \otimes \sigma)$ agrees with the Lebesgue measure on all Borel measurable sets.*

Proof. If E is a Borel subset of S^{d-1}, then

$$\mu((x,y]) = \int_{(x,y]} d\mu = \int_x^y r^{d-1} dr = \frac{y^d - x^d}{d}$$

we conclude that

$$\pi_*(\mu \otimes \sigma)(\pi((x,y] \times E)) = \mu((x,y])\sigma(E) = (y^d - x^d)|S_E|$$

Now $\pi((x,y] \times E) = yS_E - xS_E$ is Borel measurable, because S_E is Borel, and by scaling properties of subsets of $\mathbf{R}^d$, we conclude that

$$|\pi((x,y] \times E)| = |yS_E| - |xS_E| = (y^d - x^d)|S_E|$$

so $\pi_*(\mu \otimes \sigma)$ agrees with the Lebesgue measure on the π system of sets of the form $\pi((x,y] \times E)$, where E is Borel. But this family contains a countable basis of $\mathbf{R}^d - \{0\}$, so the π system generates all Borel subsets of $\mathbf{R}^d$, and therefore $\pi_*(\mu \otimes \sigma)$ agrees with the Lebesgue measure on all Borel subsets. □

If E is a Lebesgue measurable subset of $\mathbf{R}^d$, there are Borel subsets $A \subset E \subset B$ with $|B - A| = 0$, which implies by the above theorem that

$$(\mu \otimes \sigma)(\pi^{-1}(B - A)) = (\mu \otimes \sigma)(\pi^{-1}(B) - \pi^{-1}(A)) = 0$$

and $\pi^{-1}(A) \subset \pi^{-1}(E) \subset \pi^{-1}(B)$, so $\pi^{-1}(E)$ is measurable with respect to the completion of the product of the Borel algebra on $(0,\infty)$ with the Borel algebra on S^{d-1}, which is similarily the completion of the product algebra formed from Lebesgue measurable subsets of $(0,\infty)$ and surface measurable subsets of S^{d-1}. The extension of the measure $\pi_*(\mu \otimes \sigma)$ then agrees with all Lebesgue measurable subsets of $\mathbf{R}^d$.

Theorem 4.12. *If $f : \mathbf{R}^d \to \mathbf{R}$ is Borel measurable, then the slice $f^x(r) = f(rx)$ is Borel measurable for all $x \in S^{n-1}$, the function*

$$x \mapsto \int_0^\infty f(rx) r^{d-1} dr$$

is measurable on S^{d-1}, and

$$\int_{\mathbf{R}^d} f(x) dx = \int_{S^{d-1}} \int_0^\infty f(rx) r^{d-1} dr$$

If f is Lebesgue measurable, then f^x is Lebesgue measurable for almost all x, and the integral identity still holds.

Proof. The map $f : \mathbf{R}^d - \{0\} \to \mathbf{R}$ is also Borel measurable, and the fact that f^x is measurable for all $x \in S^{n-1}$ follows because $f^x = f \circ \pi^x$ is the composition of measurable functions, and π is Borel measurable, and we now calculate that if f is integrable, then Fubini's theorem gives

$$\begin{aligned}
\int_{\mathbf{R}^d} f(x)dx = \int_{\mathbf{R}^d-\{0\}} f(x)dx = \int_{\mathbf{R}^d-\{0\}} f(x)d\pi_*(\mu\otimes\sigma)(x) \\
= \int_{(0,\infty)\times S^{d-1}} (f\circ\pi)(r,x)d(\mu\otimes\sigma)(r,x) \\
= \int_{S^{d-1}} \int_0^\infty f(rx)d\mu(r)d\sigma(x) \\
= \int_{S^{d-1}} \int_0^\infty f(rx)r^{d-1}drd\sigma(x)
\end{aligned}$$

If f is only Lebesgue measurable, then $f\circ\pi$ is Lebesgue measurable when we *complete* the σ algebra on $(0,\infty)\times S^{d-1}$, and so we conclude f^x is Lebesgue measurable for almost all x. Similar results hold for f^r. □

Chapter 5

Integration

Lemma 5.1 (Scheffe)**.** *If $f, f_1, f_2, \dots \in L^1$, and $f_k \to f$ pointwise almost everywhere, then*

$$\int |f_n - f| \to 0 \quad \textit{iff} \int |f_n| \to \int |f|$$

Proof. First, assume f and f_n are positive functions. If we write

$$|f_n - f| = \max(f_n, f) - \min(f_n, f)$$

then $\min(f_n, f) \leqslant f$, and $\min(f_n, f) \to f$ almost everywhere, so the dominated convergence theorem guarantees that

$$\int \min(f_n, f) \to \int f$$

We can also write $\max(f_n, f) = f + f_n - \min(f_n, f)$. Thus

$$\lim_{n\to\infty} \max(f_n, f) = \lim_{n\to\infty} \int f_n$$

Then $\int |f_n - f| = 0$ if and only if $\int \max(f_n, f) = f$, which is equivalent to saying $\int f_n = f$. In general, if $\int |f_n| \to \int |f|$, we can apply Fatou's lemma to conclude that

$$\limsup \int f_n^- \leqslant \int f^- \qquad \limsup \int f_n^+ \leqslant \int f^+$$

and

$$\int f^- + \int f^+ = \liminf \int |f_n| \leqslant \liminf \int f_n^- + \liminf \int f_n^+$$
$$\leqslant \limsup \int f_n^- + \limsup \int f_n^+ \leqslant \int f^- + \int f^+$$

and this shows the liminf and limsup of $\int f_n^-$ and $\int f_n^+$ must be equal to one another, and that $\lim \int f_n^+ = \int f^+$, $\lim \int f_n^- = \int f^-$. But now applying the theorem for positive functions implies that both $\int |f^+ - f_n^+| \to 0$ and $\int |f^- - f_n^-| \to 0$, and

$$|f - f_n| = |f^+ - f_n^+ + f_n^- - f^-| \leqslant |f^+ - f_n^+| + |f^- - f_n^-|$$

and so $\int |f - f_n| \to 0$. On the other hand, if $\int |f - f_n| \to 0$, then the inequality $|f - f_n| \geqslant |f| - |f_n|$ guarantees that

$$\int |f| - \liminf \int |f_n| = \limsup \int |f| - |f_n| \leqslant 0$$

and the inequality $|f - f_n| \geqslant |f_n| - |f|$ guarantees that

$$\limsup \int |f_n| - \int |f| = \limsup \int |f_n| - |f| \leqslant 0$$

Hence the required limit of $\int |f_n|$ exists. □

Chapter 6

Interpolation

Interpolation is a core part of the 'hard' style of analysis, crunchy quantitative estimates that give strict bounds on quantitities. One basic tool here is interpolation, which 'in essense' enables one to take two results A_0 and A_1, and via combining them together obtain a family of results' between' the two results, of the form A_t for $0 \leqslant t \leqslant 1$.

The most basic example occurs in the theory of real numbers. Suppose $0 \leqslant A_0 \leqslant B_0$ and $0 \leqslant A_1 \leqslant B_1$. If we define $A_t = A_0^{1-t} A_1^t$, and $B_t = B_0^{1-t} B_1^t$, then it is trivial to verify that $A_t \leqslant B_t$, for the power function $x \mapsto x^a$ is order preserving for $a > 0$. For $t = 1/2$, we obtain the geometric mean inequality

$$\sqrt{A_0 A_1} \leqslant \sqrt{B_0 B_1}$$

Another way to see that the bound is trivial is to notice that the points $\log A_t$ are just the straight line from $\log A_0$ to $\log A_1$, and the result is established geometrically once we notice that log is order preserving.

6.1 Interpolation of Scalars

Let's consider some examples. If $A_0 = AX^{1/p}$, and $A_1 = AX^{1/q}$, then $A_t = AX^{(1-t)/p+t/q}$. These deductions are trivial, but we can still learn about the general inequality from them. For instance, if $A_0 = A_1$, then we find a lower bound $A \leqslant B_t$ over all $0 \leqslant t \leqslant 1$, and this bound can be refined to

$$A \leqslant B_t \min(B_0/B_1, B_1/B_0)^{\varepsilon}$$

for $\varepsilon < \min(t, 1-t)$. This tells us that the bound $A \leqslant B_t$ can only be sharp when B_0 and B_1 are roughly equal to one another. If $B_0 = 2^n B_1$, then we can improve the standard bound by a factor of $2^{-|n|\varepsilon}$. Since $2^{-|n|\varepsilon}$ is absolutely summable, it is a good heuristic to imply that the needed interpolation bound is negligable for $|n| \gg 0$.

The inequality $A_t \leqslant B_t$ can be easily generalized to the case where the A_t are defined in such a way that $t \mapsto \log A_t$ is a convex function (we say that the A_t are *log convex*), and $t \mapsto \log B_t$ is concave (though we normally always assume the $B_t are constant$). Thus one can interpolate upper bounds for log convex functions to obtain upper bounds across the domain. However, we cannot use interpolation to lower bound log-concave functions, nor can we extrapolate bounds from interior points (bounding A_0 and $A_{1/2}$ give us no info about A_1), but upper bounding A_0 and lower bounding A_t do gives us a lower bound on A_1. This is just the contraposition of the interpolation inequality.

Application of interpolation relies on the existence of a large class of log convex functions. If f and g are log convex, then fg is log convex because the sum of two convex functions is convex. Similarily,

6.2 Interpolation of functions

If we consider a step function $f = A\chi_Y$, then $\|f\|_p = A\mu(Y)^{1/p}$. Notice this is a log convex function of p, because for any $C > 0$, $\log(C)/p$

Chapter 7

The Hausdorff Measure

The expression of geometric properties of subsets of $\mathbf{R}^d$ requires more than can be expressed using the Lebesgue measure. For instance, curves and surfaces all have measure zero in two and three dimensions respectively, and thus we cannot distinguish them by the Lebesgue measure from any of the other nasty Lebesgue measurable subsets of measure zero. Hausdorff showed that there is a notion of 'dimension' of measure zero subsets of $\mathbf{R}^d$ which matches the dimension of corresponding curves and surfaces. Even more interestingly, Hausdorff's theory of dimension gives certain fractal subsets fractional dimension!

Here is the general idea. If $X = [0,1)$ is a unit interval, then $nX = [0,n)$ is the union of n disjoint translates of $[0,1)$. If we instead consider the unit square $X = [0,1) \times [0,1)$, then $nX = [0,n) \times [0,n)$ is the union of n^2 disjoint translates of $[0,1)$. If X is a unit cube, the nX is the union of n^3 disjoint translates of $[0,1)$, and so on and so forth. Thus, it makes sense to define the dimension of X to be the value α such that nX is the union of n^α disjoint copies of X. Note that if X is the Cantor set, then $3X$ is the union of two translates of X, so we would have to be willing to say that the Cantor set 'has dimension $\log_3 2 = 0.6309\dots$