3 Interesting Sets for the Lebesgue Measure

3.1 Invariance of Lebesgue Measure

Within $\mathbb{R}^n$ there are ways we can move sets around that seem like they either shouldn't change the measure, or should change it in predictable ways.

Definition 3.1. Let $A \subseteq \mathbb{R}^n$ and let $x \in \mathbb{R}^n$. We define the *translation* of A by x to be the set

$$x + A = \{x + a : a \in A\}.$$

Now let $t \in \mathbb{R}^{>0}$. We define the *dilation* of A by t to be the set

$$tA = \{ta : a \in A\}.$$

Lemma 3.2. *Let $A \subseteq \mathbb{R}^n$, let $x \in \mathbb{R}^n$, and let $t \in \mathbb{R}^{>0}$. Then:*

- $\lambda^*(x + A) = \lambda(A)$ *and* $\lambda^*(tA) = t^n\lambda(A)$.
- $\lambda_*(x + A) = \lambda(A)$ *and* $\lambda_*(tA) = t^n\lambda(A)$.
- *If A is measurable, then $x + A$ and tA are measurable, and $\lambda(x + A) = \lambda(A)$ and $\lambda(tA) = t^n\lambda(A)$.*

Proof. We first prove the lemma for special rectangles. If $I = [a_1, b_1] \times \cdots \times [a_n, b_n]$ is a special rectangle, then

$$x + I = [a1 + x1, b1 + x1] \times \cdots \times [a_n + x_n, b_n + x_n]$$

so by definition,

$$\lambda(x + I) = \prod_{i=1}^{n}(b_i + x_i - (a_i + x_i) = \prod_{i=1}^{n}(b_i - a_i) = \lambda(I).$$

Similarly,

$$tI = [ta_1, tb_1] \times \cdots \times [ta_n, tb_n]$$

and so

$$\lambda(tI) = \prod_{i=1}^{n} t(b_i - a_i) = t^n \prod_{i=1}^{n}(b_i - a_i) = t^n\lambda(I).$$

Now we want to extend this to all Lebesgue measurable sets. But this just follows from the steps of the construction of the Lebesgue measure. Clearly the result holds for special polygons; and then the set of special polygons contained in $x + G$ or tG is the set of translations or dilations of special polygons contained in G. Thus the result holds for open sets. Similarly, the result must hold for compact sets, and thus for inner and outer measure. Finally, since the result holds for inner and outer measure, it holds for the measure of measurable sets.

□

We'd like to generalize these two operations a bit further. We want to include translations and dilations, and also some other operations like rotations.

Definition 3.3. Suppose $f : U \to V$ is a function of vector spaces. We say that f is *affine* if

$$f(ax + (1-a)y) = af(x) + (1-a)f(y)$$

for any vectors $x, y \in U$ and scalars $a \in \mathbb{R}$.

This basically tells us that we don't preserve vectors, but we do preserve lines: a point on the line from x to y gets mapped to a point on the line from $f(x)$ to $f(y)$.

Exercise 3.4. *Prove that $f : U \to V$ is affine if and only if there is a linear function $L : U \to V$ and a vector $v \in V$ such that $f(x) = v + L(x)$ for every $x \in U$. Further, this choice of L and v is unique.*

An affine transformation combines a translation and a linear function, but we already understand translations. So let's see what linear functions do to the Lebesgue measure. We wish to prove the following statement:

Theorem 3.5. *Let T be a $n \times n$ matrix, and let $A \subseteq \mathbb{R}^n$. Then*

$$\lambda^*(TA) = |\det T|\lambda^*(A)$$
$$\lambda_*(TA) = |\det T|\lambda_*(A)$$

Further, if A is measurable, then TA is measurable, and

$$\lambda(TA) = |\det T|\lambda(A)$$

sketch. We'll specialize to just proving this for an open set G; once that's proven, we can extend it to the rest of measurable sets. And we can cover G by small cells that we've already understood.

So let $J = [0,1) \times \cdots \times [0,1)$. This is not a special rectangle, but it is a rectangle. Clearly $\lambda(J) = 1$. Then since T is continuous, we can see that $T(J)$ must be measurable. We set $\rho = \frac{\lambda(TJ)}{\lambda(J)} = \lambda(TJ)$. And we claim that $\lambda(TA) = \rho\lambda(A)$.

From here we're essentially going to tile G from the inside with copies of J. We can divide $\mathbb{R}^n$ into translated copies of J of the form $[a_1, a_1+1) \times (a_2, a_2+1) \times \cdots \times [a_n, a_n+1)$. Take all the ones that are inside G. Then tile the remainder with $1/2 \times 1/2 \times \cdots \times 1/2$ rectangles, and then 1/4, and so on. By following this process we can write $G = \bigcup_{k=1}^\infty J_k$; each J_k is disjoint, and is a translation of a dilation of J.

For any rectangle $J_k = z_k + t_k J$ we see that $\lambda(J_k) = t_k^n \lambda(J)$, and thus

$$\lambda(TJ_k) = \lambda(Tz_k + t_k TJ) = \lambda(t_k TJ) = t_k^n \lambda(TJ) = \rho\lambda(J_k).$$

Then we can see that

$$\begin{aligned}
\lambda(TG) &= \lambda\left(\bigcup_{k=1}^\infty TJ_k\right) \\
&= \sum_{k=1}^\infty \lambda(TJ_k) = \sum_{k=1}^\infty \rho\lambda(J_k) \\
&= \rho \sum_{k=1}^\infty \lambda(J_k) = \rho\lambda\left(\bigcup_{k=1}^\infty J_k\right) = \rho\lambda(G).
\end{aligned}$$

This proves our formula for open sets; by our sort of standard Lebesgue construction, we can extend this to any Lebesgue measurable set.

To prove the theorem, we have to prove that $\rho = |\det T|$. We can just say this is a theorem of linear algebra: the determinant of a matrix is the volume of the image of the unit cube. But if we want to prove it, we can follow this outline:

If T is invertible, then it's a theorem of linear algebra that T can be written as a product of "elementary" matrices, which correspond to the three row operations. We can show that this result holds for any elementary matrix; since the determinant is multiplicative, that implies that it holds for any invertible matrix.

If T is multiplying one dimension by a scalar, then (without loss of generality) $T(J) = [0, c) \times [0,1) \times \cdots \times [0,1)$, so $\det(T) = c$ and $\lambda(TJ) = |c|$. If T is a row-switching matrix, then $\det T = 1$ and $TJ = J$ so $\lambda(TJ) = 1$.

If T is a row-addition matrix, then $\det T = 1$. Showing that $\lambda(TJ) = 1$ is a bit trickier. But we can carefully choose a set

$$A = \{-cx_2 \leq x_1 \leq 0, 0 \leq x_i \leq 1 \text{ for } i > 1\}$$

and then if we apply the row-adding matrix

$$T = \begin{bmatrix} 1 & c & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

then $T(A)$ is just A reflected across the first coordinate. Thus $\lambda(TA) = \lambda(A)$. Since we know that $\lambda(TA) = \rho\lambda(A)$ that proves that $\rho = 1 = \det T$.

Conversely, if T is invertible, hen the determinant of T is zero, so we want to show that $\rho = 0$, or equivalently, that $\lambda(TA) = 0$. It's sufficient to show that $T\mathbb{R}^n$ has zero measure. But since $\det T = 0$, we know that the kernel is non-trivial, and by the rank-nullity theorem $\dim T(\mathbb{R}^n) < \dim \mathbb{R}^n$. We proved that any proper affine subspace has measure zero, and thus $T(\mathbb{R}^n)$ has measure zero.

(Technically we only proved this if the affine subspace is a special rectangle, but there's nothing really interesting about proving it for the rotated versions.)

□

We'll finish this discussion by mentioning a particularly important class of affine transformations:

Definition 3.6. Suppose V is an inner product space. We say a linear transformation $L : V \to V$ is *orthogonal* if $\langle L(u), L(v)\rangle = \langle u, v\rangle$.

We say a $n \times n$ matrix A is *orthogonal* if A is invertible and $A^{-1} = A^T$ the transpose of A.

Exercise 3.7. *Prove that a matrix is orthogonal if and only if the associated linear transformation is orthogonal.*

Exercise 3.8. *Prove that if L is orthogonal, then $|\det L| = 1$. Hint: use theorem 3.5 and use $A = B(0, 1)$.*

This shows that if L is an orthogonal matrix, then $\lambda(A) = \lambda(LA)$ for any measurable set A; that is, orthogonal matrices preserve measure. Since translations also preserve measure, we can generalize just a hair further.

Definition 3.9. Let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ such that there is a $z \in \mathbb{R}^n$ and an orthogonal matrix L such that $\Phi(x) = z + Lx$ for any $x \in \mathbb{R}^n$. Then we say that Φ is a *rigid motion*. Notice that Φ is an affine transformation.

Remark 3.10. The set of rigid motions on $\mathbb{R}^n$ form a group, known as the *Euclidean group* or the group of rigid motions.

It is equivalent to ask that Φ be an *isometry*, that is, that Φ preserve distances: we say that Φ is an isometry if

$$|\Phi(x) - \Phi(y)| = |x - y|$$

for any $x, y \in \mathbb{R}^n$.

3.2 A non-measurable set

In this section we will construct (after a fashion) a set $E \subseteq \mathbb{R}^n$ that is not measurable.

We begin by looking at the set $\mathbb{Q}^n \subset \mathbb{R}^n$. For any fixed $x \in \mathbb{R}^n$ we can consider the set of translations $x + \mathbb{Q}^n$. It's easy to see that $y \in x + \mathbb{Q}^n$ if and only if $y - x \in \mathbb{Q}^n$.

Exercise 3.11. *Prove that the translates of $\mathbb{Q}^n$ partition $\mathbb{R}^n$. That is, if $x, y \in \mathbb{R}^n$, then either $x + \mathbb{Q}^n = y + \mathbb{Q}^n$ or $(x + \mathbb{Q}^n) \cap (y + \mathbb{Q}^n) = \varnothing$.*

We sometimes might call these translates "cosets" of $\mathbb{Q}^n$.

It's clear that $\mathbb{R}^n = \bigcup_{x \in \mathbb{R}^n} x + \mathbb{Q}^n$. But each set on the right occurs infinitely many times. If we assume the axiom of choice, we can pick exactly one $x \in \mathbb{R}^n$ in each translate of $\mathbb{Q}^n$; let $E \subset \mathbb{R}^n$ be the set of these chosen points. Then $\mathbb{R}^n = \bigcup_{x \in E}(x + \mathbb{Q}^n)$, and this union is disjoint.

But we can also turn this statement around! For every $x \in \mathbb{R}^n$, we have exactly one $y \in E$ such that $x - y \in \mathbb{Q}^n$. But if we write $x - y = z$, we see that there is exactly one $z \in \mathbb{Q}^n$ such that $x - z = y \in E$. So instead we can write a disjoint union

$$\mathbb{R}^n = \bigcup_{z \in \mathbb{Q}^n} z + E.$$

And this union is disjoint.

But this by itself generates a problem. It's easy to see from this that if E is measurable, it must have positive measure. For

$$\lambda^*(\mathbb{R}^n) = \lambda^* \left(\bigcup_{z \in \mathbb{Q}^n} z + E \right) \leq \sum_{z \in \mathbb{Q}^n} \lambda(z + E) = \sum_{z \in \mathbb{Q}^n} \lambda(E).$$

If $\lambda^*(E) = 0$, then, we have $\lambda^*(\mathbb{R}^n) = 0$ which is clearly false.

But we will show that $\lambda_*(E) = 0$. Let K be any compact subset of E; we will show that $\lambda(K) = 0$. Fix $D = B_1(0) \cap \mathbb{Q}^n$ to be the rational points in the unit ball. Then D is a bounded, countably infinite set. We know that

$$\bigcup_{r\in D} r+K \subseteq \bigcup_{r\in D} r+E$$

is a countably infinite disjoint union. We compute that

$$\lambda\left(\bigcup_{r\in D} r+K\right) = \sum_{r\in D}\lambda(r+K)\sum_{r\in D}\lambda(K).$$

If $\lambda(K)>0$, then this sum is infinite; but since $DandK$ are bounded, the union is bounded and thus has finite measure. Thus we must have $\lambda(K)=0$. Since this holds for any compact set $K\subseteq E$, this implies that $\lambda_*(E)=0$.

Then $0=\lambda_*(E)<\lambda^*(E)$, and so E is not measurable.

Corollary 3.12. *If $A\subseteq\mathbb{R}^n$ is measurable and $\lambda(A)>0$ then there is a non-measurable subset $B\subseteq A$.*

Proof. Let E be the set we constructed above; then we can write

$$A=\bigcup_{x\in\mathbb{Q}^n}\left((x+E)\cap A\right).$$

Since A has positive measure, and this is a countable union, there is at least one $x_0\in\mathbb{Q}^n$ such that $(x_0+E)\cap A$ has positive outer measure. Then set $B=(x_0+E)\cap A$. By our argument from above, $\lambda_*(B)=0$, but $\lambda^*(B)>0$. Thus $B\notin\mathcal{L}$. □

This same logic, with some care, can be used to generate important paradoxical results.

Fact 3.13 (Banach-Tarski). *Let A,B be any two bounded subsets of $\mathbb{R}^3$ with nonempty interior. Then we can write both sets as finite disjoint unions $A=\bigcup_{k=1}^n A_k$, $B=\bigcup_{k=1}^n B_k$, and define rigid motions $\Phi_k:\mathbb{R}^3\to\mathbb{R}^3$, such that $\Phi_k(A_k)=B_k$.*

This is "paradoxical" because A and B need not have the same measure, but we know the rigid motions Φ_k preserve measure. In the famous example, we take A to be a ball of radius 1, and B to be the disjoint union of two balls of radius 1. Though an explicit construction that uses the axiom of choice, Banach and Tarski showed that you can divide A into five disjoint pieces, and use rigid motions of each piece to produce B.

However, there is no rigid motion such that $\Phi(A)=B$.

Exercise 3.14. *Prove that there are disjoint subsets $A,B\subseteq\mathbb{R}^n$ such that*

$$\lambda^*(A\cup B)<\lambda^*(A)+\lambda^*(B)$$
$$\lambda_*(A\cup B)>\lambda_*(A)+\lambda_*(B).$$

Exercise 3.15. *Let $A, B, C \subseteq \mathbb{R}^n$ such that $A \subseteq C$ and $\lambda(B \cap C) = 0$. Then A, B are not necessarily disjoint but they are separated in a measure theoretic sense. Prove that $\lambda^*(A \cup B) = \lambda^*(A) + \lambda^*(B)$.*

3.3 Cantor Sets and Lebesgue Functions

In this section we're mostly going to stay in $\mathbb{R}$, although there are perfectly reasonable generalizations to $\mathbb{R}^n$ and we'll try to mention them.

We've already seen the Cantor set $C \subseteq \mathbb{R}$ in section 2.1.4. We removed a union of open middle thirds, and saw what was left. Here we can generalize this.

Choose a sequence of positive real number l_k such that $1 > 2l_1 > 4l_2 > \cdots > 2^k l_k > \ldots$. We can start with the closed interval $[0, 1]$ and remove an open interval from the middle of length $1-2l_1$, leaving $[0, l_1] \cup [1-l_1, 1]$ as the remainder. We denote the middle open interval $(l_1, 1 - l_1) = J_{1/2}$.

From each of these closed intervals we can remove a middle bit of length $l_1 - 2l_2$, leaving four intervals of length l_2. We call the removed intervals $J_{1/4} = (l_2, l_1 - l_2)$ and $J_{3/4} = (1 - l_1 + l_2, 1 - l_2)$.

At the kth step of this process, we have remaining 2^k intervals of length l_k, and have removed $2^k - 1$ intervals which we have labeled $J_{i/2^k}$ for $1 \leq i \leq 2^k - 1$.

Let us denote the limiting set

$$A = [0, 1] \setminus \bigcup_{k \in \mathbb{N}, 1 \leq i \leq 2^k - 1} J_{i/2^k}.$$

A is the complement of a union of open intervals, and thus A is closed and hence compact. We see that $\lambda(A) = \lim_{k\to\infty} 2^k l_k$.

We obtain the original Cantor set C by taking $l_k = 3^{-k}$ for each k. Then $\lambda(C) = \lim_{k\to\infty}(2/3)^k = 0$.

The generalized Cantor sets have one more interesting property: they are *nowhere dense*.

Definition 3.16. A set A is *nowhere dense* if its interior is empty. That is, A is nowhere dense if $A^\circ = \varnothing$. Consequently, $\overline{A^\circ} = \varnothing$ as well.

Why is A nowhere dense? if A has non-empty interior, then it must contain an open interval I with positive length r. We can choose a k such that $2^{-k} \leq r$, and then A is contained in a union of disjoint intervals of length 2^{-k}. Thus $I \not\subset A$.

You might think that this implies that A has zero measure. Recall we used the original Cantor set to show you can have an uncountable set with zero measure. But we can build "fat" Cantor sets with positive measure. In fact, if we set

$$l_k = \frac{\theta k + 1}{(k+1)2^k}$$

then $\lambda(A) = \theta$. This works for any $\theta \in [0, 1)$. Thus we can have a nowhere dense set of positive measure, and in fact of just about as much measure as we like.

We can now define the *Lebesgue function* associated to A. We'll set $J_0 = (-\infty, 0)$ and $J_1 = (1, \infty)$. Then it's easy to define a function $f : A^C \to [0, 1]$ by $f(x) = r$ for every $x \in J_r$. We know that the interval J_r is entirely to the left of J_s if $r < s$, so f is an increasing function.

Further, f is continuous on A^C. Informally, we can convince ourselves of this because it seems like the function must be locally constant. But there are infinitely many infinitely small sub-intervals, so it's possible something weird is going on.

However, suppose $|x - y| < l_k$. Then if $x \in J_r$ and $y \in J_s$, one of two things must happen. One possibility is that $r = s$, in which case $f(x) = f(y)$. But if $r \neq s$, then the intervals J_r and J_s must be relatively small, and close together. Both r and s will have to have denominators $\geq 2^{-k}$, and thus $|f(x) - f(y)| < 2^{-k}$. Thus f must be continuous, and in fact uniformly continuous.

(You can see Jones p. 88 for a careful proof of this last fact, but it's mostly some careful work with this definition as a limit).

Thus f is continuous on A^C. It turns out that we can extend f to be continuous on the closure of A^C—which is in fact all of $\mathbb{R}$.

Exercise 3.17. *Let $E \subseteq \mathbb{R}^n$ and $f : E \to \mathbb{R}$ be uniformly continuous. Then there is a unique function $F : \overline{E} \to \mathbb{R}$ such that F is continuous and $F(x) = f(x)$ for all $x \in E$.*

In our particular case we will call this extension the *Lebesgue function* corresponding to A. It is a continuous non-decreasing function $f : \mathbb{R} \to [0, 1]$ that has the property that $f(x) = r$ for any $x \in J_r$. By the intermediate value theorem, it is surjective onto $[0, 1]$.

This function is also an almost-bijection between the extended Cantor set A and the open interval $(0, 1)$. First, if $x < y$ then $f(x) < f(y)$, unless $x, y \in \overline{J_r}$ for some r. In particular, if $x, y \in A$ then $f(x) < f(y)$ unless the open interval (x, y) is one of the J_r. Thus x is *almost* strictly increasing on A.

In particular, f is strictly increasing on A except there are two points outputting $i/2^k$ for each i, k. So let $B = \{\inf(J_r)\} \cup \{0\}$ be the set of all the left endpoints of the intervals J_r. Then $f : (A \setminus B) \to (0, 1)$ is strictly increasing surjective function, and thus a bijection.

By standard set theory/cardinality arguments, this means that A has the same cardinality as $(0, 1)$.

Exercise 3.18. *If f is the Lebesgue function associated to some Cantor set A, then $f(1 - x) + f(x) = 1$ for any x.*

3.4 Non-Borel Measurable Sets

In this section we will prove that not every measurable subset of $\mathbb{R}$ is Borel. When we talk about product measures, we'll extend this result to $\mathbb{R}^n$.

Let C be the ternary Cantor set, and let f be the Lebesgue function associated to it. f is strictly increasing on C, but not on $\mathbb{R}$; but we can make it strictly increasing by defining $g(x) = x + f(x)$. Since f is continuous and nondecreasing, g is continuous and strictly increasing. Then g gives us a homeomorphism from $[0, 1]$ onto $[0, 2]$.

We first claim that $g(C)$ has positive measure. Since g is a bijection, we know that

$$g(C) = [0, 2] \setminus g(C^C) = [0, 2] \setminus g\left(\bigcup J_r\right) = [0, 2] \setminus \bigcup g(J_r).$$

But on J_r the function f is constant, so the function g is just given by $g(x) = x + r$. Thus g maps each open interval J to another open interval of the same length, and so $\lambda(g(J_r)) = \lambda(J_r)$.

Then we have

$$\lambda \bigcup g(J_r) = \sum \lambda g(J_r) = \sum \lambda(J_r) = 1$$

since we worked this out when we studied the Cantor set. Thus we have

$$\lambda(g(C)) = \lambda([0, 2]) - \lambda(g(C^C)) = 2 - 1 = 1.$$

So g has already done something strange: it's a homeomorphism between a set of measure zero and a set of measure 1. Somehow it stretches the volume of C infinitely.

But now let's consider this set $g(C)$. It's a closed set of measure 1. And since it has positive measure, by corollary 3.12, there is some set $B \subseteq g(C)$ that is not measurable. Then we define $A = g^{-1}(B)$.

We know that $A \subseteq C$, and thus $\lambda^*(A) \leq \lambda(C) = 0$. Thus A is measurable because the Lebesgue measure is complete. So we just have to prove that A is not Borel. But since g is a homeomorphism, A is Borel if and only if $g(A) = B$ is Borel (by corollary 2.44). But B is not measureable, and so it's definitely not Borel. Thus A isn't Borel either.

So we've constructed a measure zero set which isn't Borel, but is measurable (because it's measure zero). We can easily build a positive measure set that's measurable but not Borel by, like, taking $A \cup [5, 7]$. This will have measure 2, but still not be Borel.

One more observation to make here: we know homeomorphisms preserve Borel sets. But they clearly *don't* preserve measurable sets, since A is measurable and $g(A) = B$ is not.