Exercise Class 1

The Motivation behind σ-Algebras

These notes largely follow the first chapter of [Tao11].

The concept of the measure generalizes the notion of the old ideas of length, area and volume of solid bodies. These old ideas are rather basic and can be easily explained to a layman. One might then wonder why we need to introduce new (rather abstract) definitions such as algebras, σ-algebras, Borel sets, etc. In this class, we will put our focus towards the question:

What are σ-algebras and why are they important?

The goal is to obtain an intuitive understanding of why the notion of σ-algebra is needed and why the definition as presented below is the right one. To that end, we briefly revisit the classical methods used to compute the measure (i.e. the length, the area, or the volume) of a set.

1 Limitations of the Classical Approach

The first mathematical approach to computing the measure of a body consisted of decomposing it into (finitely many) components, rearranging them by rigid motions (e.g. translations or rotations), and then reassembling those components to form a simpler body for which the measure is assumed to be known. The measure is presumably unchanged after performing these operations. For general curved figures often only lower and upper bounds could be determined by computing the measures of inscribed and circumscribed bodies (e.g. Archimedes' computation of the area of the circle). That is, to obtain the "true measure" we would need to pass to the limit.

These approaches are justified by geometrical or physical intuition. Once we try to define a measure using an analytical foundation, however, it is no longer apparent how a measure should be defined. The intuition of definining the measure of a body to be the sum of the measures of its components runs into an immediate problem. Typically, a solid body consists of an (uncountably) infnite number of points, all of which have measure zero. The product $\infty \cdot 0$ is indeterminate. Another issue is that sets with the same number of points need not have the same measure. Think of the intervals $[0, 1]$ and $[0, 2]$. They have the same cardinality (using the bijection $x \mapsto 2x$), but $[0, 2]$ is twice as long as $[0, 1]$. So theoretically, we can decompose $[0, 1]$ into disjoint components (points in this case) and reassemble them to form the interval $[0, 2]$. Of course, one can point out that the issue arises from the fact that the number of components is uncountably infinite. However, even if we restrict ourselves to finitely many components, one runs into trouble.

Theorem 1 (Banach–Tarski paradox)**.** *Given a solid ball B in 3-dimensional space, there exists a decomposition of the ball into a finite number of disjoint subsets (in fact, just five would suffice), which can then be reassembled in a different way to yield two identical copies of the original ball.*

The problem in the paradox is that the components are highly pathological. Their construction even requires the *axiom of choice*[1].

[1]The Banach-Tarski paradox does not occur in a set theory model without the axiom of choice. This, however, makes for a nasty set-theoretic universe. If you want to learn more about the Banach–Tarski paradox, see https://youtu.be/s86-Z-CbaHA.

Apparently, there seems to be some kind of ambiguity concerning the measure of certain sets. Do these "counterexamples" matter in practice? No. Are they of importance for the underlying theory? Very much so.

2 σ-Algebras

σ-algebras are built to be the domain of measures. A priori, it is not obvious that we cannot make the domain of the measure to be the power set of the space in question. The naive belief might be that we can assign a measure to every set; and on finite or countably infinite spaces the naive belief is fine. Once we work over an uncountably infinite space, we run into trouble for a number of reasons, the most famous of which is the Banach–Tarski paradox. Another famous example are the so-called *Vitali sets*. These are sets to which we cannot properly assign a notion of measure (with the properties of the Lebesgue measure). Do we throw away the Lebesgue measure? No. Instead, we accept that our measure only works on some sets and we then consider this smaller collection of privileged subsets. These subsets will be called *measurable sets*. As it turns out, the collection of these sets form a σ-algebra (see Theorem 5).

Let X be some set, and let $\mathcal{P}(X)$ denote its power set.

Definition 2. The set system $\mathcal{A} \subseteq \mathcal{P}(X)$ is called a *σ-algebra* (or *σ-field*) if it satisfies the following properties:

i) $X \in \mathcal{A}$;

ii) closed under complementation: $A \in \mathcal{A} \implies A^c \in \mathcal{A}$;

iii) closed under countable unions: $A_1, A_2, \dots \in \mathcal{A} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Remark. It follows by applying De Morgan's law that a σ-algebra is also closed under countable intersections.

The definition of a σ-algebra may seem odd by itself. However, the reasons for the definition being the way it is becomes apparent as we go along.

Let $\mathcal{A} \subseteq \mathcal{P}(X)$. Our goal is to define a measure $\mu : \mathcal{A} \to [0, \infty]$ with properties we deem reasonable (e.g. translation-invariance and rotation-invariance). We could choose $\mathcal{A} = \mathcal{P}(X)$. However, as mentioned before, for some sets there is no sensible notion for their measure and we quickly run into trouble.

Thus given that we cannot take $\mathcal{A}$ to be the power set of X, what kind of structure must we enforce on $\mathcal{A}$ to obtain a useful notion of measure? Preferably, we hope such a collection to be as big as possible so that we can properly assign measure to as many subsets as possible. To see why Definition 2 is the right structure, let us walk through each of the properties and interpret them in terms of measures. So let $\mathcal{A}$ be a σ-algebra as in Definition 2. Then

i) X must be in $\mathcal{A}$

This is because we would like to know the measure of the entire space.

ii) If A is in $\mathcal{A}$, then A^c is in $\mathcal{A}$.

If A and X are in $\mathcal{A}$, then we know $\mu(A)$ and $\mu(X)$. So it is sensible to define $\mu(X \setminus A) = \mu(X) - \mu(A)$. Hence, this requirement is a reflection of a desired algebraic relationship for measures. In the language of probability theory, this becomes even more reasonable. There, our measure is a probability measure and the measure of a set A denotes the probability that the outcome is in A. If we can ask whether the outcome is in A, we also want to be able to ask whether the outcome is *not* in A. Thus we insist that $A^c \in \mathcal{A}$.

iii) $\mathcal{A}$ is closed under countable unions.

We can imagine why finite unions are important. If we have finitely many disjoint sets A_i, then we ought to be able to define $\mu(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$. Again, this is a reflection of an algebraic requirement of measures on σ-algebras. This also agrees with our geometric intuition of decomposing a body and summing up the measures of the components.

The extension to countable unions is more subtle but it is probably the most important structural element of a σ-algebra. Let us recall how limits of sets were defined.

Definition 3. For a sequence of sets $A_1, A_2, \dots \subseteq X$ we define

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m, \qquad \limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m.$$

If, in fact, $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$, we say that the limit $\lim_{n\to\infty} A_n$ exists.

Since we want to do analysis and take limits, we will also want countable unions of measurable sets to be measurable. If we only allow finite unions, then our theory of measures will end up being incompatible (albeit, not completely) with limiting operations. This extension is what ends up making Lebesgue's theory of integration so useful, as we suddenly have a theory of integration that plays well with limits (unlike the Riemann integral).

Conclusion: We want to define the notion of measure for various sets of a space X. Ideally, every set should be assigned a measure, however, that turns out to be unrealistic in practice. Thus we decide that we are only able to measure nice enough sets, which we call measurable. "Weird" sets should not have a measure.

We have yet to define measurability of a set. Here, *Carathéodory's criterion of measurability* comes in handy.

Definition 4 (Carathéodory)**.** A set $A \subseteq X$ is called *measurable* if $\forall B \subseteq X$

$$\mu(B) = \mu(B \cap A) + \mu(B \setminus A).$$

With this definition, things fall into place quite nicely. One can now proof that the collection of measurable sets forms a σ-algebra.

Theorem 5. *Let $\mu : \mathcal{P}(X) \to [0, \infty]$ be a measure. Then*

$$\Sigma := \{A \subseteq X : A \text{ measurable}\}$$

is a σ-algebra.

3 An Application in Probability Theory

In probability theory, the notion of σ-algebra has an intuitive interpretation. Here, the σ-algebra is used to characterize information that can be observed, i.e. the σ-algebra contains information. A simple example illustrates this idea.

Example 6. Consider a fair coin flip. The possible outcomes are heads H or tails T. Thus the sample space is $X = \{H, T\}$. The possible events are that nothing can happen, something can happen or it comes up heads/tails. Recall that the σ-algebra is supposed to be the domain of the (probability) measure. For each event, we know how to compute the proability. Hence these events then form the σ-algebra. That is, $\mathcal{A} = \{\varnothing, X, \{H\}, \{T\}\}$. In this case, every subset of X is measurable, i.e. $\mathcal{A} = \mathcal{P}(X)$.

Now let us modify this example. Assume now, that we do not know whether the coin is fair. The outcomes are still $X = \{H, T\}$. However, the σ-algebra is different now. We do not know what the probability for H or T is, i.e., we cannot assign a measure to the events $\{H\}$ and $\{T\}$. Thus they are not in the σ-algebra anymore. In this case, we obtain the trivial σ-algebra $\mathcal{A} = \{\varnothing, X\}$; a remarkably useless amount of information. Different σ-algebras thus correspond to a different "amount of knoweldge".

Let us finally consider another, slightly more complicated example.

Example 7. Imagine betting on a game that involves flipping a coin repeatedly and observing whether it comes up heads (H) or tails (T). Assuming that you are infinitely wealthy, the game can go on forever. The sample space Ω must consist of all possible outcomes, i.e.

$$\Omega = \{H, T\}^{\infty}.$$

After n flips of the coin, you might want to revise your betting strategy in advance of the next flip. The observed information can be described in terms of the 2^n possibilities for the first n flips. Formally, this is encoded as the σ-algebra

$$\mathcal{G}_n = \{A \times \{H, T\}^{\infty} : A \subseteq \{H, T\}^n\}.$$

That is, $\mathcal{G}_n$ contains the observed information after n flips of the coin. Note that $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \ldots \mathcal{G}_{\infty}$. Such an increasing sequence of σ-algebras is called a *filtration* and is a frequently used object in probability theory.

Sources

[Tao11] Terence Tao. *An introduction to measure theory*. Vol. 126. 2011. URL: https://terrytao.files.wordpress.com/2011/01/measure-book1.pdf.