Why Do We Need to Care About σ-Algebras?

Richard Chapling v1.2 5 March 2020

1. Definitions

For reference, we recall

Definition 1 (σ-algebra). Let Ω be a set. A set $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ (i.e., a set of subsets of Ω) is called a *σ-algebra* if the following three properties hold:

F1 $\Omega \in \mathcal{F}$.

F2 If $A \in \mathcal{F}$, $A^c \in \mathcal{F}$ as well.

F3 If $\{A_k\}_{k=1}^{\infty} \in \mathcal{F}$ is a countable collection of subsets, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ as well.

Definition 2 (Probability measure). Given $(\Omega, \mathcal{F})$ as above, a function $\mathbb{P} \colon \mathcal{F} \to [0,1]$ is called a *probability measure* if

P1 $\mathbb{P}(\Omega) = 1$

P2 If $\{A_k\}_{k=1}^{\infty} \subseteq \mathcal{F}$ is a countable collection of disjoint sets, then $\mathbb{P}\big(\bigcup_{i=1}^{\infty} A_k\big) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$.

In this course, Ω is called the *sample space*. The pair $(\Omega, \mathcal{F})$ is usually called a *measurable space*. The elements of $\mathcal{F}$ are called *events*. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space*.[1]

2. The good news: countable sample spaces

The good news is that if Ω is countable, there are no problems with taking $\mathcal{F} = \mathcal{P}(\Omega)$. Indeed, it is easy to show that if the singleton sets $\{\omega\}$ are all in $\mathcal{F}$, then every $A \subseteq \Omega$ is in $\mathcal{F}$. (Exercise: do this.)

(Notice that for once, nothing goes wrong in the extension from finite to infinite: this is because our definitions use countable unions rather than finite ones.)

3. The bad news: uncountable sample spaces

Continuous probability distributions normally use $\mathbb{R}$ (or $\mathbb{R}^d$) as their sample space. Now it is no longer true that $\mathcal{P}(\Omega)$ can be generated by only singletons, because with only countable unions and complements, we can't form uncountable sets with uncountable complements from these, for example. Worse, there are serious problems with trying to assign probabilities to all subsets of $\mathbb{R}$, as we shall now see. It is easier to use $[0,1)$ as a sample space, since we can equip it with a uniform probability measure, $\mathbb{P}$, say.

We define a relation on $[0,1)$ by $r \sim s$ if and only if $r - s \in \mathbb{Q}$. It is easy to check that this is an equivalence relation. Use the Axiom of Choice to pick one element of $[0,1)$ in each equivalence class, and define V to be the set of these chosen elements. V is called a *Vitali set*.

For any subset $A \subseteq [0,1)$ and $r \in \mathbb{R}$, define

$$A_r \coloneqq \{a + r \mod 1 : a \in A\},$$

where $\mathrm{mod}\,1$ means to take the fractional part. The probability measure is uniform, so we must have $\mathbb{P}(A_r) = \mathbb{P}(A)$: writing $s = r \bmod 1$, we have $A = ((A \cap [0, 1-s))) \cup (A \cap [1-s, 1))$, and then

$$A_r = A_s = ((A \cap [0, 1-s)) + s) \cup ((A \cap [1-s, 1)) + s - 1),$$

[1] In more general measure theory, elements of $\mathcal{F}$ are called *measurable sets*. Instead of $\mathbb{P} \colon \mathcal{F} \to [0,1]$ with $\mathbb{P}(\Omega) = 1$, a more general measure $\mu \colon \mathcal{F} \to [0,\infty]$ uses $\mu(\emptyset) = 0$. $(\Omega, \mathcal{F}, \mu)$ is called a *measure space*.

so A_r is just A cut into two pieces that are then translated by different amounts.

Now, $[0,1)$ is a countable union of such sets: if we enumerate the rationals in $[0,1)$ as $(q_k)_{k=1}^{\infty}$, then

$$[0,1) = \bigcup_{k=1}^{\infty} V_{q_k},$$

because for each $\alpha \in V$, $\{\alpha + q_k \bmod 1 : k \in \mathbb{N}\}$ fills the equivalence class represented by α. Moreover, the union is disjoint, since the if the real numbers α and β are in distinct equivalence classes, so too are $\alpha + q_i$ and $\beta + q_j$. Therefore we should have

$$1 \overset{\text{P1}}{=} \mathbb{P}([0,1)) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} V_{q_k}\right) \overset{\text{P2}}{=} \sum_{k=1}^{\infty} \mathbb{P}(V_{q_k}) = \sum_{k=1}^{\infty} \mathbb{P}(V).$$

But this is impossible: either $\mathbb{P}(V) = 0$, so the sum is 0, or $\mathbb{P}(V) > 0$, so the sum diverges. Either way, it cannot be equal to 1, so we obtain a contradiction. Since we really want to have a uniform probability measure, *we have to reject the idea that V can be assigned a measure.*

You may hope that this problem arises from only using countable unions, and that we could fix this problem by replacing these by uncountable unions, but (Exercise) one can show that if $a_i \geqslant 0$, $\sum_{i \in I} a_i$ can be finite only if $\{i \in I : a_i \neq 0\}$ is countable, which shows that even if we allow our sums to be uncountable, we don't actually improve the theory. The σ-algebra is the usual mathematical answer to these problems: we restrict the subsets of Ω we can talk about so that what we say about them makes sense.

A. Appendix: Some comments on the use of the Axiom of Choice here

You may be aware that the Axiom of Choice is somewhat controversial, and therefore might be concerned by its use here. It is actually necessary to use it to construct non-measurable sets: if we eschew AC as an axiom, there *are* models of $\mathbb{R}$ in which every set is measurable. These are called *Solovay models.*

You may be inclined to use such a model instead, since then we no longer have to worry about σ-algebras, but there are two things to bear in mind:

1. To construct such a model, one has to accept a different axiom, about the existence of an "inaccessible" cardinal. You haven't done any serious set theory yet, so you probably don't really have an opinion about this, but suffice to say that such axioms are also the subject of debate, at least amongst those who are interested in such things.

2. However, one can show that such a model of $\mathbb{R}$ has some *very weird* properties: for example, it is possible to partition $\mathbb{R}$ into more subsets than it has elements. Do you really want to be working in a model that does things like this?

This being the case, most mathematicians prefer to stick with there being non-measurable sets, since we have ways around this (and this is hardly the strangest thing about the standard version of $\mathbb{R}$ anyway: $\mathbb{R}$ is actually an extremely complicated object, much more so than even most mathematicians realise).

Finally, you are surely aware of the *Banach–Tarski paradox*, namely that it is possible to decompose a sphere into a finite number of pieces and reassemble it, using rotations and translations, into two spheres of the same size as the original.[2] Since this operation does not preserve the volume (!), the pieces have to be non-measurable sets in $\mathbb{R}^3$. Usually this is seen as an example to support the position "The Axiom of Choice has weird consequences, so let's not use it", but this is not really correct: the mistake is really in thinking of geometrical sets as made of points in the first place. $\mathbb{R}$ and $\mathbb{R}^d$ as we normally think of them have more structure than a mere set of points (by having a topology, for example, or a way of measuring distances): it is not enormously surprising that constructions that do not take this extra structure into account should be discordant with our intuition. So it is too for Vitali sets: because they are not *useful*, and they conflict with our intuition, we are not really bothered if our theory cannot deal with them (however they are constructed).

[2] What is an anagram of BANACHTARSKI? BANACHTARSKIBANACHTARSKI.