Statistics C205A Lecture 1 Notes

Daniel Raban

August 26, 2021

1 Motivation for Measure Theory

1.1 Desired properties of probability

Here are a few experiments. We want a way to compute the "probability" of certain outcomes.

Example 1.1. Suppose we toss a coin, with outcomes $\{0, 1\}$. If the coin is fair, $\mathbb{P}(\{0\}) = \mathbb{P}(\{1\}) = 1/2$.

Example 1.2. Choose a point, uniformly randomly from $[0, 1]$, and call the outcome U. We want to be able to calculate:

$$\mathbb{P}(U \in [0, 1/2]) = 1/2.$$

What about $\mathbb{P}(U = 0.5)$? Intuitively, this should equal 0 because each point should have the same probability as the others; but if any point had positive probability, since there are infinitely many points, there would be infinite total probability, which is impossible.

Now let $\mathbb{Q}$ denote the rational numbers. What is $\mathbb{P}(U \in \mathbb{Q})$? This should also be 0 because each point has probability 0. Here, we are making the assumption of **countable additivity**, which says if the events $E_1, E_2, \dots$ are disjoint, then $\mathbb{P}(E_1 \cup E_2 \cup \cdots) = \sum_{i=1}^\infty \mathbb{P}(E_i)$.

Now what is $\mathbb{P}(U \text{ is irrational})$? This is 1, because we have total probability 1 and because we have ruled out the rational case with probability 0.

What about $\mathbb{P}([1/3, 2/3])$. You may guess that this equals $1/3$, which is the "length" of the interval. This equals $\mathbb{P}([2/3, 1])$, as well; that is, we want **translation invariance**, the property that $\mathbb{P}(A) = \mathbb{P}(A + x)$ for any A, x such that $A + x \subseteq [0, 1]$.

1.2 Constructing a non-measurable set

We want to define a function $\mathbb{P} : 2^{[0,1]} \to [0, 1]$ which measures the probability of any subset of $[0, 1]$, has countable additivity, and has translation invariance. Unfortunately, there is no way to satisfy all these properties. We will show an example of a set $A \subseteq [0,1]$ which will break these properties. To avoid the issue of sets getting translated outside of the interval, we will work on the unit circle S^1, rather than the unit interval. That is, we want a set $A \subseteq S^1$ such that $\mathbb{P}(A)$ does not make sense.

Example 1.3. Take any $x_0 \in S^1$ and rotate it an angle of 1 (radian) to get x_1. Do it again to get $x_2, x_3, \ldots$. Then, for any $x, y \in S^1$, say that $x \sim y$ if we may reach x from y by finitely many jumps (clockwise or counterclockwise). This relation is *transitive*, i.e. if $x \sim y$ and $y \sim z$, then $x \sim z$.

Let us denote the equivalence class of x by E_x. If we let $x+i$ denote the point obtained after i jumps (with $i \in \mathbb{Z}$), then $x+i \neq x+j$ for $i \neq j$, lest $i-j$ be an integer multiple of 2π. So $E_x = \{x+i : i \in \mathbb{Z}\}$. For $x, y \in S^1$, either $E_x = E_y$ ($x \sim y$) or $E_x \cap E_y = \varnothing$. This gives a partition of S^1 into equivalence classes. Because the equivalence classes are all countable and S^1 is uncountable, we must have uncountably many equivalence classes E_x.

Now suppose we can construct our set $A \subseteq S^1$ such that $S^1 = \bigcup_{i \in \mathbb{Z}} A + i$ is a disjoint union. Using countable additivity and translation invariance,

$$\underbrace{\mathbb{P}(S^1)}_{=1} = \sum_{i \in \mathbb{Z}} \mathbb{P}(A+i) = \sum_{i \in \mathbb{Z}} \mathbb{P}(A).$$

So there is no way to define $\mathbb{P}(A)$ to make this make sense. All that is left is to exhibit such a set A. Form the set A by choosing 1 representative from each equivalence class.[1]

To address this, we will work with a suitable subclass of 2^{S^1} that will suffice for most of our applications.

In one and two dimensions, one does need to have countable additivity to construct such a counterexample. In 3 dimensions and higher, there is the Banach-Tarski paradox:

Theorem 1.1 (Banach-Tarski). *The sphere S^3 admits a decomposition into disjoint pieces $A_1, \ldots, A_8$ such that A_1, A_2, A_3, A_4 can be rotated and translated to form a copy of S^3 and A_5, A_6, A_7, A_8 can be rotated and translated to form a copy of S^3.*

1.3 σ-fields

In general, we will have a set Ω and define a subset $\Sigma \subseteq 2^\Omega$ called a σ-field. This will be the collection of all sets which it makes sense to assign probability.

Definition 1.1. A **σ-field** is a nonempty collection $\Sigma \subseteq 2^\Omega$ such that

1. $A \in \Sigma \implies A^c \in \Sigma$
2. $A_1, A_2, \dots \in \Sigma \implies \bigcup_{i=1}^\infty A_i \in \Sigma$.

[1] Implicitly, we are using the axiom of choice to perform this operation.

Example 1.4. $\{\varnothing, \Omega\}$ is a σ-field.

Note that Σ always contains $\varnothing$ and Ω.