Setup of a Probability Space

Narayana Santhanam

May 25, 2023

This material complements Section 1.2 of the text.

Now that we know a little bit about sets, we can lay the foundation for probability. In common parlance, people understand probability from a relative frequency (or betting odds) perspective. The approach we take (in these notes and the text) is more formal.

A probability model contains three elements:

- Ω, a set, called the Sample Space or universe. There are no formal restrictions on Ω, but you have to choose one that will be useful for the problem or application you have. To be useful, the elements of Ω need to be a mutually exclusive, exhaustive set of outcomes of the experiment in question.

- $\mathcal{E}$, a set of subsets of Ω. Every element of $\mathcal{E}$ is a set, that is for all $E \in \mathcal{E}$, $E \subset \Omega$. Equivalently $\mathcal{E}$ is a subset of the power set of Ω, namely $\mathcal{E} \subset 2^{\Omega}$. We call the elements of $\mathcal{E}$, *events*.

  So $\mathcal{E}$ is the set of all possible events, and only these get probabilities. It does NOT mean that elements not in $\mathcal{E}$ get zero probability—the elements not in $\mathcal{E}$ are simply excluded from consideration altogether. Note that $\mathcal{E}$ can be different from the power set of Ω, namely $\mathcal{E}$ can exclude certain subsets of Ω. So when we say "A is an event", it has the following implication A is a subset of Ω, but not just any subset—it also belongs to the exclusive club $\mathcal{E}$.

  This level of subtlety is unnecessary when Ω is countable, and in this class, when Ω is countable, we take $\mathcal{E} = 2^{\Omega}$ (so every subset of Ω gets to be an event). But when Ω is continuous (uncountable), in all applications we consider, $\mathcal{E}$ is very different from the power set of Ω. This distinction means that we handle discrete and continuous cases very differently.

$\mathcal{E}$ cannot be an arbitrary subset of the power set 2^{Ω}. It must contain Ω, be closed under "countable" unions, intersections and complements. A countable union means you can index the sets involved in the union by natural numbers (but potentially infinitely many sets could participate). The technical term for $\mathcal{E}$ is *sigma-algebra*, but the name is not important at this introductory level (and this is the last we will speak of it).

- $\mathbb{P} : \mathcal{E} \to [0, 1]$, a function, called a *probability law* or probability assignment that assigns to each $E \in \mathcal{E}$, a number $\mathbb{P}(E)$ that represents the probability of the event E. The assignment also cannot be arbitrary, it needs to satisfy the axioms of probability.

Note that the text does not mention $\mathcal{E}$ hoping that you will not ask too many questions. This is not uncommon in beginning courses in probability, particularly in engineering. But it also makes many finer points confusing in the later part of the course, so it is better to be aware of what is happening.

The axioms of probability are the following. First note that since $\mathbb{P} : \mathcal{E} \to [0, 1]$, every event is assigned a probability ≥ 0 and ≤ 1. This is automatic from how we define the function $\mathbb{P}$, but sometimes (as in your text), people make it explicit and write this as an axiom by itself.

1. Recall Ω is always guaranteed to be in $\mathcal{E}$. We must assign $\mathbb{P}(\Omega) = 1$.

2. If A and B are disjoint events (that is, $A, B \in \mathcal{E}$ and $A \cap B = \{\}$, then we must assign

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

   This restriction applies not just to finite disjoint unions, but countable disjoint unions—if $A_1, A_2, \ldots$ are a sequence of disjoint events (meaning for all n, $A_n \in \mathcal{E}$ and if $i \neq j$, then $A_i \cap A_j = \{\}$), then

$$\mathbb{P}(\cup_n A_n) = \sum_n \mathbb{P}(A_n).$$

Specifically, if A is an event, point 2 above immediately implies $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

Event space

Say we have a sample space Ω. Now we have to decide what the events are, that is what is $\mathcal{E}$. To take the next step, we have to first determine if Ω is discrete (countable) or continuous (uncountable). We will not cover formal definitions here, but go by examples (since we will not encounter very funny cases anyway).

A set is "countable" informally means you can write the elements of the set one after another on an infinite sheet of paper given infinite time. For example, all natural numbers (but also all integers, though it is slightly trickier to see). A set being uncountable means that even given the infinite sheet of paper and infinite time, you could not write all the elements of the set one after another.

Countability, as you can see, is a much more sophisticated concept than just "finite".

Countable and uncountable sets

Finite sets are always countable, as are sets such as $\mathbb{N}$ (the set of all natural numbers), and $\mathbb{Z}$ (the set of all integers). So is the set of all $k-$tuples whose entries come from any countable set, say all $k-$tuples of natural numbers (which, as noted before, we call a $k-$fold Cartesian product of natural numbers, writing it as $\mathbb{N}^k$, or $\underbrace{\mathbb{N} \times \cdots \mathbb{N}}_{k \text{ terms}}$). Less obvious examples are the set of all rational numbers.

A (somewhat less obvious) NON-example is a countable Cartesian product of a countable set $\underbrace{\mathbb{N} \times \cdots \mathbb{N}}_{\infty \text{ terms}}$—such a set is uncountable (or as we say, continuous).

The set of all real numbers $\mathbb{R}$ is uncountable, so is any interval $[a, b]$, $(a, b]$, $[a, b)$, (a, b) (if $a < b$). These are the only uncountable (or continuous) sets we pay attention to in the course. But for those interested, the Cantor set is uncountable as well. The power set of $\mathbb{N}$ (the set of all subsets of $\mathbb{N}$) is uncountable as well.

What is $\mathcal{E}$?

Discrete When Ω is countable, we will take $\mathcal{E}$ to be the power set of Ω, 2^Ω (the set of all subsets of Ω). This means that when Ω is countable, all subsets of Ω are allowed to participate as events, and they can be assigned probabilities (you can assign 0 probability, but they get an assignment).

Continuous When the sample space Ω is uncountable, we primarily consider the case where the sample space is $\mathbb{R}$ (the set of all real numbers). Here every interval $[a, b]$, $(a, b]$, $[a, b)$, (a, b), for all $a \leq b$, (a and b are allowed to take values $\pm\infty$ as well) are events. Note that $a = b$ is allowed, so the event $[a, a]$ is simply a set with one element, $\{a\}$. These events all have to be assigned probabilities consistent with the axioms of probabilities.

While intervals are the only events we focus on, the set of events $\mathcal{E}$ is much more complicated. Because the set of events $\mathcal{E}$ must be closed under countable unions, $\mathcal{E}$ contains very complicated sets, including for example, the complement of Cantor's set. Because it must be closed under complements, it must contain Cantor's set too. And many even more bizzare sets. We just don't focus on them, but the mechanisms we use apply to those sets as well.

There will also be subsets of Ω that we exclude to come up with a probability law satisfying the axioms of probability. Once again, the excluded sets are not about what we focus or not, nor about whether they have non-zero probability. On the contrary, assigning any number to them breaks the axioms of probability irredeemably. Such sets are called non-measurable sets, and are a lot of fun to consider, and the experiences in this module contains an example of such sets. Also associated with non-measurable sets is an amazing "paradox" called the Banach-Tarski paradox, also added under experiences.

Probability assignment

Discrete The example in class or the text shows you a natural way to handle the discrete case. Here the set of all events is 2^Ω. You simply assign every singleton sets $\{\omega\}$ (for every element $\omega \in \Omega$) a number $\mathbb{P}(\{\omega\}) = p_\omega$ between 0 and 1, such that

$$\sum_\omega p_\omega = 1.$$

The numbers p_ω need NOT be the same for all $\omega \in \Omega$, as our example in class shows. The assignment $\mathbb{P}(\{\omega\}) = p_\omega$ is called a *probability mass function.*

After this, the probability axiom corresponding to disjoint unions takes care of all other events. Convince yourself that all axioms are automatically satisfied if you assign any other event $E \subset \Omega$ the probability

$$\mathbb{P}(E) = \sum_{e \in E} p_e.$$

Continuous In the continuous case, the event space is lot more complicated. To see this, suppose you wanted to figure out a uniform probability law over the interval [0,1]. Intuitively, this means that the probability of any interval must be proportional to its length.

We could not replicate this intuition by assigning probabilities to individual singleton events of $[0, 1]$ as in the discrete case. Recall that a singleton event is an event with one element, $\{\frac{1}{2}\}$ for example. By our intuition, because we are aiming to construct in this case a uniform probability law, all singleton events must be assigned the same value. Now our problems start—a non-zero value implies that any non-trivial interval, say $[0, \frac{1}{2}]$ or $[0, 1]$, will end up with an uncountably infinite number of disjoint singleton events (and will have a subset with countably infinite number of disjoint events), all with non-zero probabilities. So any non-zero probability is out for singletons. The only real option is to assign every singleton event probability zero.

But there is a more serious issue. In the discrete case, we said any other event has its probability determined by the countable-disjoint union axiom. But here, events have uncountably many disjoint components. So the axioms are unhelpful in assigning an event like $[0, \frac{1}{2}]$. We need to do something else.

Instead, we take a non-negative function $f : [0, 1] \to \mathbb{R}^+$ (here $\mathbb{R}^+$ means the set of all non-negative real numbers, includes 0). We now say that the probability of an event A is the integral of the function in A,

$$\mathbb{P}(A) = \int_{x \in A} f(x) dx.$$

While you may have learnt the integral as an analogy to summation, integration is most definitely not summation. The above expression is not to be confused with the discrete analogy where we add probabilities. Similarly, $f(x)$ is NOT the probability of x. It is simply a mechanism to assign probabilities—and you get probabilities only via an integration.

What if $A = [a, a]$, a singleton event $\{a\}$? Well for non-pathalogical choices of f,

$$\mathbb{P}([a, a]) = \int_a^a f(x)dx = 0.$$

So, we have assigned singleton events 0 as expected. But the probability of other events, for example $[0, \frac{1}{2}]$ is

$$\mathbb{P}([0, \frac{1}{2}]) = \int_0^{\frac{1}{2}} f(x)dx.$$

Such a function f is called a *probability density function* or pdf for short. What f reflects a uniform law on [0,1]? Obviously

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{else.} \end{cases}$$

If you have been paying attention, we have to allow sets like the Cantor set $\mathcal{C}$ to be an event. So what is

$$\int_{\mathcal{C}} f(x)dx?$$

In Reimann integral formulations, we draw a blank. But the integral we use here is a generalization of Reimann integration, called the Lebesgue integral. Long story short, $\int_{\mathcal{C}} f(x)dx$ are well defined (and evaluates to 0). But for simple intervals and the simple events we will consider, the Reimann and Lebesgue integrals coincide, so this is the last we will use the term Lebesgue integration.

To emphasize, in the continuous case, the disjoint-union axiom of probability does not help us specify the probability law from the probability assignment on singleton events. Instead, we have an all-new integration approach to assign probabilities to events. In the continuous case, you get probabilities only via integration.