Math 525: Lecture 1

January 5, 2018

1 Motivation

Example 1.1. A coin is tossed whose sides are labelled H and T.

Arguably, if we knew the exact conditions of the coin toss, we could use the laws of physics to accurately predict its outcome. Of course, this approach is impractical! It is more useful to assign a belief to each outcome in order to aid our prediction-making. For example, we may assign a $1/2$ "probability" to each outcome if we believe the coin to be fair. But what do we mean by "probability", and what laws (a.k.a. axioms) does it abide by?

Remark 1.2. The coin tossing example is a mundane one: probability has much more exciting applications (e.g., financial analysis, machine learning, etc.).

2 Probability space

2.1 Sample space

Recall Example 1.1. H and T are referred to as *outcomes* of the coin tossing *experiment*. We can gather these outcomes into a single set $\Omega = \{H, T\}$, referred to as the *sample space*.

Example 2.1. A coin is tossed 3 times. We record the result of each coin toss. The sample space for this experiment is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Note that $|\Omega| = 2^3$. If the coin is tossed n times, $|\Omega| = 2^n$.

The example above deals with a finite sample space. Can we think of an example with an *infinite* sample space?

Example 2.2. A coin is tossed repeatedly until the first occurrence of H, at which point we record the total number of tosses. The sample space is

$$\Omega = \{1, 2, \ldots\} = \mathbb{N}.$$

The example above deals with a countably infinite sample space. Can we think of an example with an *uncountable* sample space?

Example 2.3. A coin is tossed. Once the result (H or T) is recorded, the coin is tossed again. This process continues ad infinitum. The sample space is

$$\Omega = \{(\omega_1, \omega_2, \ldots) \colon \omega_i = H \text{ or } \omega_i = T\}.$$

For example, the outcome in which all tosses result in heads is written $(H, H, H, \ldots) \in \Omega$ (or, omitting the parentheses and commas for brevity, $HHH\ldots$).

2.2 σ-algebra

Recall Example 2.1, in which a coin is tossed 3 times. We may be interested in answering questions of the following form:

1. What is the "probability" that there are an even number of heads?
2. What is the "probability" that the first toss is a head?
3. What is the "probability" that the first toss is **not** a head?
4. What is the "probability" that the first toss is **not** a head **or** there are an even number of heads?

The set of outcomes in which there are an even number of heads is a subset of the sample space Ω. Primarily, it is the set

$$A_1 = \{HHT, HTH, THH, TTT\}.$$

Similarly, the set of outcomes in which the first toss is a head is

$$A_2 = \{HHH, HHT, HTH, HTT\}.$$

We can continue in this way to describe the set of outcomes for the remaining two questions above. They are $A_3 = A_2^c \equiv \Omega \setminus A_2$ and $A_4 = A_1 \cup A_3$.

Definition 2.4. A subset of Ω is called an *event*.

As hinted at above, we would like to be able to take complements of an event A to describe its non-occurrence and unions of events A and B in order to describe two (or more) events occurring simultaneously.

Definition 2.5. The power set of Ω, denoted 2^Ω, is the set of all subsets of Ω.

Example 2.6. Let $\Omega = \{1, 2, 3, \ldots, 6\}$, corresponding to rolling a six-sided dice. Then,

$$2^\Omega = \{\emptyset, \{1\}, \{2\}, \ldots, \{6\}, \{1,2\}, \{1,3\}, \ldots, \{5,6\}, \ldots, \Omega\}.$$

Note that $|2^\Omega| = 2^{|\Omega|}$ (in this case, $2^{|\Omega|} = 2^6 = 64$).

Definition 2.7. An *algebra* (a.k.a. field) on Ω is a set $\mathcal{F} \subset 2^\Omega$ satisfying the following properties:

1. $\emptyset \in \mathcal{F}$.

2. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

3. if $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$.

Proposition 2.8. *Let $\mathcal{F}$ be an algebra on Ω.*

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

2. *If $A_1, \ldots, A_n \in \mathcal{F}$, then $A_1 \cup \cdots \cup A_n \in \mathcal{F}$.*

3. *If $A_1, \ldots, A_n \in \mathcal{F}$, then $A_1 \cap \cdots \cap A_n \in \mathcal{F}$.*

Proof. The first claim is established by noting that $\Omega = \emptyset^c$. The next claim is established by writing

$$A_1 \cup \cdots \cup A_n = (A_1 \cup \cdots \cup A_{n-1}) \cup A_n$$

and applying induction. The last claim follows by De Morgan's law:

$$A_1 \cap \cdots \cap A_n = ((A_1 \cap \cdots \cap A_n)^c)^c = (A_1^c \cup \cdots \cup A_n^c)^c .$$

□

Algebras are fine for finite sample spaces, but they are not particularly useful for infinite sample spaces. When dealing with infinite sample spaces, we require the notion of a σ-algebra:

Definition 2.9. Let $\mathcal{F}$ be an algebra on Ω such that $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$ whenever $A_1, A_2, \ldots \in \mathcal{F}$. We call $\mathcal{F}$ a *σ-algebra* (a.k.a. σ-field).

Example 2.10. For any sample space Ω, $\mathcal{F} = \{\emptyset, \Omega\}$ is the so-called *trivial* σ-algebra.

Example 2.11. For any sample space Ω, $\mathcal{F} = 2^\Omega$ is the so-called *discrete* σ-algebra.

Remark 2.12. At this point, you may ask, why do we need σ-algebras aside from the discrete σ-algebra? After all, the discrete σ-algebra has every possible event that we might be interested in. As it turns out, we can always use the discrete σ-algebra for finite (and even countably infinite) sample spaces, but attempting to do so for an uncountable sample space turns out to have some bad consequences, which are out of the scope of this course.

Exercise 2.13. Let $\mathcal{F}$ be a σ-algebra on Ω. If $A_1, A_2, \ldots \in \mathcal{F}$, then $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$ and $A_1 \cap A_2 \cap \cdots \in \mathcal{F}$.

Exercise 2.14. Let $\{\mathcal{F}_\alpha\}_{\alpha \in \mathcal{A}}$ be a family of σ-algebras (on Ω). Then, $\mathcal{F} = \cap_{\alpha \in \mathcal{A}} \mathcal{F}_\alpha$ is a σ-algebra.

2.3 Probability measure

To each event A, we would like to associate a number $0 \le \mathbb{P}(A) \le 1$, the *probability* of the event A occurring. The intuitive interpretation of $\mathbb{P}(A) = 0$ (resp. $\mathbb{P}(A) = 1$) is that the event A *does not occur with certainty* (resp. *occurs with certainty*). A larger value of $\mathbb{P}(A)$ means that we "believe" A to be more likely to occur. We define $\mathbb{P}$ rigorously below:

Definition 2.15. Given a σ-algebra $\mathcal{F}$ on Ω, a *probability measure* is a function $\mathbb{P}\colon \mathcal{F} \to [0,1]$ satisfying the following properties:

1. $\mathbb{P}(\emptyset) = 0$.
2. $\mathbb{P}(\Omega) = 1$.
3. $\mathbb{P}$ is *countably additive*. That is, if $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint (i.e., $A_i \cap A_j = \emptyset$ whenever $i \ne j$), then
$$\mathbb{P}\left(\sum_{n\ge 1} A_n\right) = \sum_{n \ge 1} \mathbb{P}(A_n).$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space.*

Example 2.16. The probability space associated with Example 1.1 is $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = \{H, T\}$, $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\}$, and

$$\mathbb{P}(\{a\}) = \begin{cases} p & \text{if } a = H \\ 1-p & \text{if } a = T \end{cases} \qquad \text{where} \qquad 0 \le p \le 1.$$

If the coin is fair, $p = 1/2$.

Remark 2.17. A probability measure is an example of a *measure* μ, which satisfies all of the above points save for $\mu(\Omega) = 1$. In this case, $(\Omega, \mathcal{F}, \mu)$ is simply called a measure space.

Exercise 2.18. A probability measure $\mathbb{P}$ satisfies the following properties:

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
2. $\mathbb{P}(A) \le \mathbb{P}(B)$ whenever $A \subset B$.
3. If $A_1, \ldots, A_n \in \mathcal{F}$, then
$$\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} \mathbb{P}(A_1 \cap \cdots \cap A_n).$$

Recall that definition of continuity by sequences: f is continuous if $f(x_n) \to f(x)$ whenever $x_n \to x$. We now establish that $\mathbb{P}$ is continuous in a similar sense:

Proposition 2.19. *Let $A_1, A_2, \ldots$ be an increasing sequence in $\mathcal{F}$ (i.e., $A_1 \subset A_2 \subset \cdots$) and define $A = \cup_{n\ge 1} A_n$. Then, $\mathbb{P}(A_n) \to \mathbb{P}(A)$. Similarly, if $B_1, B_2, \ldots$ is a decreasing sequence in $\mathcal{F}$ (i.e., $B_1 \supset B_2 \supset \cdots$), we define $B = \cap_{n \ge 1} B_n$. Then, $\mathbb{P}(B_n) \to \mathbb{P}(B)$.*

Proof. It is sufficient to prove the first statement, since the second follows by defining $A_n = B_n^c$ and noting that

$$1 - \mathbb{P}(B_n) = \mathbb{P}(B_n^c) = \mathbb{P}(A_n) \to \mathbb{P}(A) = \mathbb{P}(B^c) = 1 - \mathbb{P}(B).$$

To prove the first statement, define $C_1 = A_1$ and $C_n = A_n \setminus A_{n-1}$ for $n > 1$. Then,

$$\begin{aligned}
\mathbb{P}(A) = \mathbb{P}(\cup_{n\geq 1} C_n) = \mathbb{P}(C_1) + \sum_{n>1} \mathbb{P}(C_n) = \mathbb{P}(A_1) + \sum_{n>1} \mathbb{P}(A_n) - \mathbb{P}(A_{n-1}) \\
= \mathbb{P}(A_1) + \lim_{N\to\infty} \sum_{n=2}^{N} \mathbb{P}(A_n) - \mathbb{P}(A_{n-1}) = \lim_{N\to\infty} \left\{ \mathbb{P}(A_1) + \sum_{n=2}^{N} \mathbb{P}(A_n) - \mathbb{P}(A_{n-1}) \right\} \\
= \lim_{N\to\infty} \mathbb{P}(A_N)
\end{aligned}$$

as desired. □