Lecture 2 - Axiomatic Definition of Probability

Limits of the classical definition

- The principle of indifference is ambiguous
- Each outcome assigned the same probability
    - e.g. biased coin
- Limited to two specific spaces (finite sets or subsets of $\mathbb{R}^n$)
    - Tossing a coin until first head is observed
- Mathematical inconsistency

The problem of d'Alembert

- Consider the following game: toss two coins and win if at least one head is observed, otherwise we lose. What is the probability of winning?
- Answer: Let A be the event that the game is won
- Since $\Omega = \{HH, HT, TH, TT\}$, $A = \{HH, HT, TH\}$ and of course $P(A) = \frac{3}{4}$
- D'Alembert's answer: If the first coin comes up head, the second toss will not happen as the game is already settled. Therefore $\Omega = \{H, TH, TT\}$, $A = \{H, TH\}$, so $P(A) = \frac{2}{3}$
- 3/4 is the right answer proven experimentally
- Ambiguity is assigning equal probabilities to all outcomes!

Bertrand's Paradox

- Consider an equilateral triangle inscribed in a unit circle. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle?
- Attempt 1: Choose any point on the circle, and then the angle will be chosen at random. The sample space is $[0, 2\pi)$ and the probability is then $\frac{1}{3}$.
- Attempt 2: Consider a chord perpendicular to the radius and choose the distance from the centre of the circle. The sample space is $[0, 1]$ and the probability is then $\frac{1}{2}$.
- Attempt 3: Identify the chord by its centre point. Then each chord can be uniquely identified by all the points inside the circle. The sample space is the area of a unit circle, and the probability is then $\frac{1}{4}$.
- The source of this paradox is that in each case we used a different sample space.
- The fact that all outcomes are equally likely in one space does not necessarily mean that they are equally likely in another space!
- This, all three cases concern different random experiments!

Probabilistic Space

- Probabilistic space is a triple: $(\Omega, \mathcal{F}, P)$ where
  - Sample space - Ω
  - Collection of events (σ-algebra) - $\mathcal{F}$
  - Probability measure - P
- Sample space is exactly the same as in classical probability, but it can be infinite or even uncountable
- Events are subsets of the sample space, as before, so we can perform operations on events like union, intersection, etc.
- A collection of events $\mathcal{F}$ is a collection of subsets of Ω, which contains all possible events; so $\mathcal{F} \subseteq 2^{\Omega}$.

- We can't assume that $\mathcal{F} = 2^\Omega$, because of non-measurable sets, so for countable Ω we can assume that $\mathcal{F} = 2^\Omega$, but for uncountable Ω we can only say that $\mathcal{F} \subseteq 2^\Omega$.
- No matter what exactly $\mathcal{F}$ is, we always want to be able to apply all set-theoretic operations, or in other words, we want the outcomes of those operations to be events as well (i.e. to belong to $\mathcal{F}$). This is guaranteed is we assume that $\mathcal{F}$ is a σ-algebra.
- A collection of $\mathcal{F} \subseteq 2^\Omega$ is called a σ-algebra, if:
    - $\Omega \in \mathcal{F}$
    - If $A \in \mathcal{F}$ then $A' \in \mathcal{F}$
    - If $A_1, A_2, \ldots \in \mathcal{F}$ then $A_1 \cup A_2 \cup \ldots \in \mathcal{F}$ for any countable sum of events
- If $\mathcal{F} = 2^\Omega$ for countable Ω then $\mathcal{F}$ is a σ-algebra
- Borel Algebra: assume that $\mathcal{F}$ at least contains all events of the form: "the outcome less than a", "the outcome between a and b". Then $\mathcal{F}$ must contain all possible intervals, open or closed, finite or infinite, e.g. $[a, b), (a, b), (-\infty, a], (b, \infty)$ etc. Such a collection is called Borel σ-algebra.
- The probability measure is a real-valued function P defined on a σ-algebra $\mathcal{F} \subseteq 2^\Omega$, which satisfies:

1. Nonnegativity: $P(A) \geq 0$ for all $A \in \mathcal{F}$
2. Normalization: $P(\Omega) = 1$
3. Additivity: For any sequence of disjoint events $A_1, A_2, \ldots \in \mathcal{F}$:

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Probability over the countable sample space

- Informally: If Ω is countable, it suffices to assign a "probability value" to every outcome and then the probability of any event A is just a sum of the probability values of all outcomes which belong to A.
- Formally: Let $\Omega = \{\omega_1, \omega_2, \ldots\}$ be a countable set and let $\mathcal{F} = 2^{\Omega}$.
- To every ω_n assign a real number $p_n \geq 0$ such that

$$\sum_{n=1}^{\infty} p_n = 1$$

- The probability of any event $A \subseteq \Omega$ is defined as a sum of p_n, over all $\omega_n \in A$:

$$P(A) = \sum_{n:\omega_n \in A} p_n$$

- Therefore, $p_n = P(\{\omega_n\})$ is the probability of outcome ω_n.
- Of course, all of this also holds for a finite sample space Ω.
- For uncountable sample space we cannot assign a probability value to each outcome and compute the probabilities of events.