**MTH 224, Spring 2024**

**Instructor: Bahman Angoshtari**

**Lecture 2**

**Section 1.1:** axioms of probability, basics properties of probability, probability spaces with equally likely outcomes

## 2.1. Definition of probability

- Each event in the sample space has a **probability** of occurring. Intuitively, the probability is a quantitative measure of how likely the event is to occur. Given an experiment and one of its event $A$, $\mathbb{P}(A)$ denotes the probability that the event $A$ occurs. Think of $\mathbb{P}(A)$ as the proportion of times that the event $A$ would occur in the long run, if the experiment were to be repeated over and over again.
- Example: Assume that we toss a fair coin. What is the probability of Heads?
- The concept of probability is very intuitive. It has been around since (at least) the 17th century, originally to calculate the odds in gambling. However, the mathematical definition turned out not to be that simple! Indeed, the formal definition of probability was only introduce by Andrey Kolmogorov in the 1930's.

Definition 2.1. Let $S$ be a sample space. Let $\mathbb{P}$ be a function that assigns a number $\mathbb{P}(A)$ to "each" event $A$ of $S$. Then, $\mathbb{P}$ is a **probability function** (or **probability measure**) if the following 3 conditions hold (these condition are called **axioms of probability**):

(A1) $\mathbb{P}(S) = 1$

(A2) $0 \le \mathbb{P}(A) \le 1$, for any event $A \subseteq S$.

(A3) if $A_1, A_2, \dots$ are **disjoint events**, then $\mathbb{P}(A_1 \cup A_2 \cup \dots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots$.

In other words, if $A_i, A_j \subseteq S$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

Note that the third axiom only holds for **disjoint events**. Note, also, that the number of events can be **infinite**, but there must be **countably many** of them.

Remark. How do the Axioms (A1)-(A3) give us the "intuitive" concept of probability? In other words, how do (A1), (A2), and (A3) imply that $\mathbb{P}(A)$ is the proportion of times that the event $A$ would occur in the long run?

The short answer is that the "intuitive" concept of probability is a theorem called the **law of large number (LLN)** and is proved by an argument that only requires the three Axioms (A1), (A2), and (A3). We will not discuss LLN, since it is outside the scope of this course.

## 2.2. Basic properties of probability measure

- For any event $A$, we have that $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

- **Proof:** By definition $S = A \cup A^c$. Furthermore, $A$ and $A^c$ are disjoint (why?). Therefore, by (A1) and (A3), we have that

$$1 = \mathbb{P}(S) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) \implies \mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

- $\mathbb{P}(\emptyset) = 0$. Show this by using the previous result. Here, $\emptyset$ denotes the empty set, that is, a set with no elements.
- If $A = \{\omega_1, \ldots, \omega_n\}$, that is, if $A$ is the event containing outcomes $\omega_1, \ldots, \omega_n$, then $\mathbb{P}(A) = \mathbb{P}(\omega_1) + \cdots + \mathbb{P}(\omega_n)$. Show this by using (A3).
- **The inclusion-exclusion principle:** For any events $A$ and $B$ (that may or may not be disjoint), we have that

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

  - **Proof:** Note that $A \cap B$, $A \backslash B$, and $B \backslash A$ are disjoint events (use a Venn diagram) and

$$\begin{aligned} A &= (A\backslash B) \cup (A \cap B) \\ B &= (B\backslash A) \cup (A \cap B) \\ A \cup B &= (A\backslash B) \cup (A \cap B) \cup (B\backslash A) \end{aligned}$$

By (A3), the above equations yield that

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A\backslash B) + \mathbb{P}(A \cap B) \\ \mathbb{P}(B) &= \mathbb{P}(B\backslash A) + \mathbb{P}(A \cap B) \\ \mathbb{P}(A \cup B) &= \mathbb{P}(A\backslash B) + \mathbb{P}(A \cap B) + \mathbb{P}(B\backslash A) \end{aligned}$$

Thus, $\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$

$$\begin{aligned} &= \mathbb{P}(A\backslash B) + \mathbb{P}(A \cap B) + \mathbb{P}(B\backslash A) + \mathbb{P}(A \cap B) - \mathbb{P}(A\backslash B) - \mathbb{P}(A \cap B) - \mathbb{P}(B\backslash A) \\ &= \mathbb{P}(A \cap B) \end{aligned}$$

Finally, we obtain $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

- **The inclusion-exclusion principle for multiple events:**

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup A_3) =& \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) \\ & - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_2 \cap A_3) \\ & + \mathbb{P}(A_1 \cap A_2 \cap A_3) \end{aligned}$$

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_N) =& \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_N) \\ & - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \cdots - \mathbb{P}(A_{N-1} \cap A_N) \\ & + \mathbb{P}(A_1 \cap A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_4) + \cdots + \mathbb{P}(A_{N-2} \cap A_{N-1} \cap A_N) \\ & \cdots + (-1)^{N-1}\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_N) \end{aligned}$$

EXAMPLE 2.2. Alice has 35% chance to be admitted to college A and 90% chance to college B. There is 33% chance that she get admitted to both colleges. What is the probability that both colleges reject her?

SOLUTION. $\mathbb{P}(A^c \cap B^c) = \mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A \cup B) = 1 - (\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B))$
$= 1 - (0.35 + 0.9 - 0.33) = 0.08.$

EXAMPLE 2.3. There are three people with three different hats. If each person takes a random hat from the coat room, what is the probability that at least one person get her own hat?

SOLUTION. Let $A_i$ be the event that person $i$ gets her hat. The event that at least one person get her hat is then $A_1 \cup A_2 \cup A_3$. By the inclusion-exclusion principle,

$$\begin{aligned}\mathbb{P}(A_1 \cup A_2 \cup A_3) =& \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) \\ & - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3).\end{aligned}$$

We have $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{1}{3}$ (why?); $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_2 \cap A_3) = \frac{1}{6}$ (why?); and $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{6}$ (why?). Therefore:

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} - \frac{1}{6} - \frac{1}{6} - \frac{1}{6} + \frac{1}{6} = \frac{2}{3}.$$

## 2.3. Probability spaces with equally likely outcomes

- A common scenario is an experiment in which all outcomes are equally likely. The sample space of such experiments satisfy the following conditions.
  - The number of outcomes must be finite. $S = \{\omega_1, \omega_2, \dots, \omega_N\}$.
  - Outcomes are equally likely to happen: $\mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\}) = \cdots = \mathbb{P}(\{\omega_N\}) = \frac{1}{N}$ (Why is the last equation correct?)
  - Let $A = \{\omega_1, \omega_2, \dots, \omega_M\}$ be an event. We then have:
  $$\mathbb{P}(A) = \mathbb{P}(\{\omega_1\}) + \cdots + \mathbb{P}(\{\omega_M\}) = \frac{1}{N} + \cdots + \frac{1}{N} = \frac{M}{N}.$$
- Let $S$ be a sample space with eqully likely outcomes. Then, the probability of any event $A$ is calculated as follows
  $$\mathbb{P}(A) = \frac{|A|}{|S|}.$$
  Here, $|A|$ is the number of outcomes in $A$ and $|S|$ is the total number of outcomes. In general, $|E|$ denotes the number of elements of a set $E$.