## 1.1 Set Theory

Probability theory is built upon set theory. This is a very brief primer.

- A **set** is a collection of elements.
- We usually use capital letters (such as $A$) to refer to sets and lowercase letters (such as $x$) to refer to elements.
- $x \in A$ means "$x$ is an element of the set $A$."
- $x \notin A$ means "$x$ is not an element of the set $A$."
- The **empty set** or **null set** is the set with no elements. Notation: $\phi$ or $\{ \}$.
- The **universal set** $\Omega$ is the set of all elements (for the specific context).
- A **subset** $A$ of a set $B$ is a set consisting of some (or none or all) of the elements of $B$. Notation: $A \subset B$.
- Two sets $A$ and $B$ are **equal** if and only if $A \subset B$ and $B \subset A$.

### 1.1.1 Set Operations

- **Complement:** $A^{\mathsf{c}} = \{x : x \notin A\}$.
- **Union:** $A \cup B = \{x : x \in A \text{ or } x \in B\}$.
- **Intersection:** $A \cap B = \{x : x \in A \text{ and } x \in B\}$.
- **Set Difference:** $A - B = \{x : x \in A \text{ and } x \notin B\}$.

### 1.1.2 Other Set Concepts

- A collection of sets $A_1, \dots, A_n$ is **mutually exclusive** if $A_i \cap A_j = \phi$ for $i \neq j$.
- A collection of sets $A_1, \dots, A_n$ is **collectively exhaustive** if $A_1 \cup \cdots \cup A_n = \Omega$.
- A collection of sets $A_1, \dots, A_n$ is a **partition** if it is both mutually exclusive and collectively exhaustive.

### 1.1.3 De Morgan's Laws

$$(A \cup B)^{\mathsf{c}} = A^{\mathsf{c}} \cap B^{\mathsf{c}} \qquad \left(\bigcup_{i=1}^{n} A_i\right)^{\mathsf{c}} = \bigcap_{i=1}^{n} A_i^{\mathsf{c}}$$

$$(A \cap B)^{\mathsf{c}} = A^{\mathsf{c}} \cup B^{\mathsf{c}} \qquad \left(\bigcap_{i=1}^{n} A_i\right)^{\mathsf{c}} = \bigcup_{i=1}^{n} A_i^{\mathsf{c}}$$

## 1.2 Axiomatic Theory of Probability

We need a formal, principled method for assigning probabilities to sets. This will be especially useful as a foundation for complex probabilisitic reasoning (later in the course).

### 1.2.1 Basic Probability Model

- An **experiment** is a procedure that generates observable outcomes.
- An **outcome** is a possible observation of an experiment.
- The **sample space** $\Omega$ is the set of all possible outcomes.
- An **event** is a subset of $\Omega$: it is a set of possible outcomes.

### 1.2.2 Probability Axioms

A **probability measure** $\mathbb{P}[\cdot]$ is a function that maps events to real numbers. It must satisfy the following axioms:

1. **Non-negativity:** For any event $A$, $\mathbb{P}[A] \geq 0$.
2. **Normalization:** $\mathbb{P}[\Omega] = 1$.
3. **Additivity:** For any countable collective $A_1, A_2, \ldots$ of mutually exclusive events,

$$\mathbb{P}[A_1 \cup A_2 \cup \cdots] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \cdots \ .$$

- The next two properties follow directly from the axioms, and are useful to name explicitly:
  - **Complement:** $\mathbb{P}[A^{\mathsf{c}}] = 1 - \mathbb{P}[A]$.
  - **Inclusion-Exclusion:** $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.

## 1.3 Conditional Probability

- The **conditional probability** of event $A$ given that $B$ occurs is

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \ .$$

- For $\mathbb{P}[B] = 0$, $\mathbb{P}[A|B]$ is undefined.
- Conditional probability satisfies the probability axioms:
  - Non-negativity: For any event $A$, $\mathbb{P}[A|B] \geq 0$.
  - Normalization: $\mathbb{P}[\Omega|B] = 1$.
  - Additivity: For any countable collective $A_1, A_2, \ldots$ of mutually exclusive events,

$$\mathbb{P}[A_1 \cup A_2 \cup \cdots |B] = \mathbb{P}[A_1|B] + \mathbb{P}[A_2|B] + \cdots \ .$$

- **Multiplication Rule:** For two events $A$ and $B$, $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B|A] = \mathbb{P}[B]\,\mathbb{P}[A|B]$. For $n$ events $A_1, A_2, \ldots, A_n$,

$$\mathbb{P}\left[\bigcap_{i=1}^{n} A_i\right] = \mathbb{P}[A_1]\,\mathbb{P}[A_2|A_1]\,\mathbb{P}[A_3|A_1 \cap A_2] \cdots \mathbb{P}[A_n|A_1 \cap \cdots \cap A_{n-1}] \ .$$

- **Law of Total Probability:** For a partition $B_1, \ldots, B_n$ satisfying $\mathbb{P}[B_i] > 0$ for all $i$,

$$\mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A|B_i]\mathbb{P}[B_i] \ .$$

- **Bayes' Rule:** This is a method to "flip" conditioning:

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\,\mathbb{P}[B]}{\mathbb{P}[A]} \ .$$

Sometimes, it is useful to solve for the denominator using the total probability theorem. For a partition $B_1, \dots, B_n$ satisfying $\mathbb{P}[B_i] > 0$ for all $i$,

$$\mathbb{P}[B_j|A] = \frac{\mathbb{P}[A|B_j]\,\mathbb{P}[B_j]}{\mathbb{P}[A]} = \frac{\mathbb{P}[A|B_j]\,\mathbb{P}[B_j]}{\displaystyle\sum_{i=1}^{n} \mathbb{P}[A|B_i]\,\mathbb{P}[B_i]} \ .$$

## 1.4 Independence

- Two events $A$ and $B$ are **independent** if $\mathbb{P}[A \cap B] = \mathbb{P}[A]\,\mathbb{P}[B]$.
- Independence of $A$ and $B$ means that knowing if $A$ occurs cannot help predict whether $B$ also occurs (and vice versa).
- Events $A_1, \dots, A_n$ are **independent** if
  - All collections of $n-1$ events chosen from $A_1, \dots, A_n$ are independent.
  - $\mathbb{P}[A_1 \cap \cdots \cap A_n] = \mathbb{P}[A_1] \cdots \mathbb{P}[A_n]$
- This recursive condition can be tedious to check. However, in most cases, we will use independence as a modeling assumption.
- Independence means that no subset of the events can be used to help predict the occurrence of any other subset of events.
- If $A_1, \dots, A_n$ only satisfy $\mathbb{P}[A_i \cap A_j] = \mathbb{P}[A_i]\mathbb{P}[A_j]$ for all $i \neq j$, then we say they are **pairwise independent** (but not independent).

### 1.4.1 Conditional Independence

- The events $A$ and $B$ are **conditionally independent** given $C$ if

$$\mathbb{P}[A \cap B|C] = \mathbb{P}[A|C]\,\mathbb{P}[B|C] \ .$$

- Conditional independence means that, given $C$ occurs, knowing that $A$ occurs cannot help predict whether $B$ also occurs (and vice versa).
- Events $A_1, \dots, A_n$ are **conditionally independent** given $B$ if
  - All collections of $n-1$ events chosen from $A_1, \dots, A_n$ are conditionally independent given $B$.
  - $\mathbb{P}[A_1 \cap \cdots \cap A_n|B] = \mathbb{P}[A_1|B] \cdots \mathbb{P}[A_n|B]$
- Independence does not imply conditional independence.
- Conditional independence does not imply independence.

## 1.5 Counting

- If an experiment is composed of $m$ subexperiments and the $i^{\text{th}}$ subexperiment consists of $n_i$ outcomes (that can be freely chosen), then the total number of outcomes is $n_1 \, n_2 \, \cdots \, n_m$.
- Counting techniques are especially useful in scenarios where all outcomes are equally likely, since the probability of an event can be expressed as

$$\mathbb{P}[A] = \frac{\#\text{ outcomes in } A}{\#\text{ outcomes in } \Omega}$$

### 1.5.1 Sampling

- A **sampling problem** consists of $n$ distinguishable elements with $k$ selections to be made.
  - Selections may be made either with or without replacement.
  - The final outcome is either order dependent or order independent.

| | Order | |
|---|---|---|
| | Dependent | Independent |
| With Replacement | $n^k$ | $\binom{n+k-1}{k}$ |
| Without Replacement | $\frac{n!}{(n-k)!}$ | $\binom{n}{k}$ |