I.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: ϕ or $\{\ \}$.
- The **universal set** S 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 $B \subset A$.

I.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:** $A \cap B = \{x : x \in A \text{ and } x \in B\}$.
- **Difference:** $A - B = \{x : x \in A \text{ and } x \notin B\}$.

I.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 $A_1 \cup \cdots \cup A_n = S$.
- A collection of sets $A_1, \dots, A_n$ is a **partition** if it is both mutually exclusive and collectively exhaustive.

I.1.3 De Morgan's Laws

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

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

I.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).

I.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** S is the finest-grain, mutually exclusive, collectively exhaustive set of all possible outcomes.
- An **event** is a set of outcomes of an experiment.

I.2.2 Probability Axioms

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

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

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

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

I.2.3 Conditional Probability

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

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

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

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

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

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

- **Total Probability Theorem:** For a partition $B_1, \dots, B_n$ satisfying $\mathsf{P}[B_i] > 0$ for all i,

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

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

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

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

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

I.3 Independence

- Two events A and B are **independent** if $\mathsf{P}[A \cap B] = \mathsf{P}[A]\,\mathsf{P}[B]$.
- Events $A_1, \dots, A_n$ are **independent** if
  - All collections of $n-1$ events chosen from $A_1, \dots, A_n$ are independent.
  - $\mathsf{P}[A_1 \cap \cdots \cap A_n] = \mathsf{P}[A_1] \cdots \mathsf{P}[A_n]$
- 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 $\mathsf{P}[A_i \cap A_j] = \mathsf{P}[A_i]\mathsf{P}[A_j]$ for all $i \neq j$, then we say they are **pairwise independent** (but not independent).

I.3.1 Conditional Independence

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

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

- Conditional independence means that, given C occurs, A cannot help predict whether B also occurs (and vice versa).
- Independence does not imply conditional independence.
- Conditional independence does not imply independence.

I.4 Counting

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

$$\mathsf{P}[A] = \frac{\#\text{ outcomes in } A}{\#\text{ outcomes in } S}$$

I.4.1 Sampling

- Number of ways to make k selections out of n distinguishable elements

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

I.4.2 Partitions

- Say we have n elements that we want to divide into r groups such that the i^{th} group contains n_i elements for $i = 1, 2, \dots, r$ and each element appears in exactly one group so that $\sum_i n_i = n$. The number of ways to form such a **partition** is given by the multinomial coefficient

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}.$$

I.5 Independent Trials

- Consider an experiment consisting of multiple identical and independent subexperiments. Often called **independent trials.**
- **Binary Outcomes:** Each subexperiment is a success with probability p and a failure with probability $1-p$.

$$\mathsf{P}\big[\{k \text{ successes}\}\big] = \binom{n}{k} p^k (1-p)^{n-k}$$

- **Multiple Outcomes:** Each subexperiment has r possible outcomes $a_1, \dots, a_r$ with probabilities $p_1, \dots, p_r$.

$$\mathsf{P}\big[\{n_1 \text{ occurrences of } a_1, \dots, n_r \text{ occurrences of } a_r\}\big] = \binom{n}{n_1, n_2, \dots, n_r} p_1^{n_1} \cdots p_r^{n_r}$$