Basic Concepts

1 Set Theory

Before we venture into the depths of probability, we start our journey with a field trip on set theory.

1.1 What is a set?

Sets are collections of elements or samples. Sets are denoted by capital letters (A, B). Elements are denoted by ω_i. The set containing all elements is the **sample space** Ω. The set with no elements is called the **empty set** $\emptyset$.

Now let's discuss some notation. If we say that $A = \{\omega_1, \omega_2, \dots, \omega_n\}$, then we say that A consists of the elements $\omega_1, \omega_2, \dots, \omega_n$. $\omega_i \in A$ means that ω_i is in A, whereas $\omega_i \notin A$ means that ω_i is not in A.

1.2 Comparing sets

We can compare and manipulate sets in many ways. Let's take a look at it. We say that A is a **subset** of B (denoted by $A \subset B$) if every element in A is also in B. When (at least) one element in A is not in B, then $A \not\subset B$. If $A \subset B$ and $B \subset A$ (they consist of the same elements), then A and B are equal: $A = B$. Sets are said to be **disjoint** if they have no common elements.

We can't just add up sets. But we can take the **intersection** and the **union** of two sets. The **intersection** of A and B (denoted by $A \cap B$) consists of all elements ω_i that are in both A and B. (So $\omega_i \in A \cap B$ if $\omega_i \in A$ **and** $\omega_i \in B$.) On the other hand, the **union** of A and B (denoted by $A \cup B$) consists of all elements ω_i that are in either A or B, or both. (So $\omega_i \in A \cup B$ if $\omega_i \in A$ **or** $\omega_i \in B$.)

The **set difference** $A \setminus B$ consists of all elements that are in A, but not in B. There is also the **complement** of a set A, denoted by A^c. This consists of all elements that are not in A. Note that $A^c = \Omega \setminus A$ and $A \setminus B = A \cap B^c$.

There's one last thing we need to define. A **partition** of Ω is a collection of subsets A_i, such that

- the subsets A_i are disjoint: $A_i \cap A_j = \emptyset$ for $i \neq j$.
- the union of the subsets equals Ω: $\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$.

2 Introduction to Probability

It's time to look at probability now. Probability is all about experiments and their outcomes. What can we say about those outcomes?

2.1 Definitions

Some experiments always have the same outcome. These experiments are called **deterministic**. Other experiments, like throwing a dice, can have different outcomes. There's no way of predicting the outcome. We do, however, know that when throwing the dice many times, there is a certain **regularity** in the outcomes. That regularity is what probability is all about.

A **trial** is a single execution of an experiment. The possible **outcomes** of such an experiment are denoted by ω_i. Together, they form the **probability space** Ω. (Note the similarity with set theory!) An **event** A is a set of outcomes. So $A \subset \Omega$. We say that Ω is the **sure event** and $\emptyset$ is the **impossible event**.

Now what is the **probability** $P(A)$ of an event A? There are many definitions of probability, out of which the **axiomatic definition** is mostly used. It consists of three axioms. These axioms are rules which $P(A)$ must satisfy.

1. $P(A)$ is a nonnegative number: $P(A) \geq 0$.
2. The probability of the sure event is 1: $P(\Omega) = 1$.
3. If A and B have no outcomes in common (so if $A \cap B = \emptyset$), then the probability of $A \cup B$ equals the sum of the probabilities of A and B: $P(A \cup B) = P(A) + P(B)$.

2.2 Properties of probability

From the three axioms, many properties of the probability can be derived. Most of them are, in fact, quite logical.

- The probability of the impossible event is zero: $P(\emptyset) = 0$.
- The probability of the complement of A satisfies: $P(A^c) = 1 - P(A)$.
- If $A \subset B$, then B is equally or more likely than A: $P(A) \leq P(B)$.
- The probability of an event A is always between 0 and 1: $0 \leq P(A) \leq 1$.
- For any events A and B, there is the relation: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Using the probability, we can also say something about events. We say two events A and B are **mutually independent** if

$$P(A \cap B) = P(A)P(B). \tag{2.1}$$

Identically, a series of n events $A_1, \ldots, A_n$ are called **mutually independent** if any combination of events $A_i, A_j, \ldots, A_k$ (with $i, j, \ldots, k$ being numbers between 1 and n) satisfies

$$P(A_i \cap A_j \cap \ldots \cap A_k) = P(A_i)P(A_j)P(A_k). \tag{2.2}$$

2.3 Conditional Probability

Sometimes we already know some event B happened, and we want to know what the chances are that event A also happened. This is the **conditional probability** of A, given B, and is denoted by $P(A|B)$. It is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{2.3}$$

The conditional probability satisfies the three axioms of probability, and thus also all the other rules. However, using this conditional probability, we can derive some more rules. First, there is the **product rule**, stating that

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) \ldots P(A_n|A_{n-1} \cap \ldots \cap A_2 \cap A_1). \tag{2.4}$$

Another rule, which is actually quite important, is the **total probability rule**. Let's suppose we have a partition $B_1, \ldots, B_n$ of Ω. The total probability rule states that

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(B_i)P(A|B_i). \tag{2.5}$$

By combining this rule with the definition of conditional probability, we find another rule. This rule is called **Bayes' rule**. It says that

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}. \tag{2.6}$$