1 Probability Space

Some Basic Knowledge about Set Theory

- A set A is **finite** if it has finite elements, i.e., $A = \{a_1, a_2, \ldots, a_n\}$; otherwise, it is infinite.
- A set A is **countable** if it has a bijection to a subset of $\mathbb{N}$, i.e., it can be "listed"; otherwise it is uncountable.
- A finite set is always countable. If a set is countably infinite, then it is in the form of $A = \{a_1, a_2, \ldots, a_n, \ldots\}$. Real numbers $\mathbb{R}$ or any interval $[a, b] \subseteq \mathbb{R}$ is uncountable.
- The **power set** of A is the collection of all its subsets, i.e., $2^A = \{B : B \subseteq A\}$.

Motivating Examples

- **Example 1: Experiment with Finite Outcomes**
  - Consider a simple **experiment**: a coin is tossed twice.
  - There are four possible **outcomes**: HH, HT, TH, TT
  - We can talk about **events** like: $A =$"no H appears", $B =$"T appears at least once", C="T appears no less than H" ...
  - Each of the above events has a **probability**: $P(A) = \frac{1}{4}$, $P(B) = P(C) = \frac{3}{4}$. Actually, B and C are the same event.

- **Example 2: Experiment with Countably Infinite Outcomes**
  - Consider another **experiment**: keep tossing a coin until H appears.
  - There are countably infinite possible **outcomes**: $H, TH, TTH, \ldots$
  - We can talk about **events** like:
    $A_k =$"H appears exactly in the kth toss" or $B_k =$"H appears in the first k tosses".
  - Then we have **probability** as follows:
    $P(A_1) = \frac{1}{2}, P(A_2) = \frac{1}{4}, P(A_3) = \frac{1}{8}, \ldots, \to 0$, which gives $\sum_{k=1}^{\infty} P(A_k) = 1$; and
    $P(B_1) = \frac{1}{2}, P(B_2) = \frac{3}{4}, P(B_3) = \frac{7}{8}, \ldots, \to 1$.

- **Example 3: Experiment with Uncountable Outcomes**
  - Consider the **experiment**: randomly take a single point in a disk of unit radius.
  - There are uncountably many possible **outcomes**: all points in the unit disk.
  - We can talk about **events** like:
    $A =$"the point is on the boundary", $B =$"the distance to $\mathbf{0}$ is between 0.3 and 0.5".
  - Then we can assign **probability**: $P(A) = 0$ and $P(B) = 0.25 - 0.09 = 0.16$.

An Informal Summary

We can conclude the followings from the above example:

- The most basic thing in an experiment is the **sample space** Ω, which is the set of all outcomes. For example, $\Omega = \{HH, HT, TH, TT\}$ for Example 1, $\Omega = \{H, TH, TTH, \dots\}$ for Example 2 and $\Omega = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ for Example 3.
- An **event is a subset of the sample space** Ω, e.g., $A = \{HT, TH, TT\}$, $B = \{H, TH, TTH, TTTH\}$ or $C = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.
- An event A **occurs** if the outcome ω of the experiment is in A, i.e., $\omega \in A$.
- A **probability function** that assigns each event a probability.

What We Need for Events

- If A and B are events, then we can think $A \cup B$, $A \cap B$ and A^c as new events "A or B", "A and B" and "no A", respectively.
- Events A and B are called *disjoint* if $A \cap B = \emptyset$.
- The empty set $\emptyset$ is the *impossible event* and the set Ω is the *certain event*.
- However, not an arbitrary subset of the sample space needs to be an event; we only want consider those events in whose occurrences we may be interested. In fact, we may have problem if we do not define events carefully!
- Let $\mathcal{F}$ be the collection of all events. Apparently we have $\mathcal{F} \subseteq 2^\Omega$, but what else do we need? The answer is **σ-field** defined as follows:

Definition: σ-Field

Let $\mathcal{F} \subseteq 2^\Omega$ be a set of subsets of Ω. We say that $\mathcal{F}$ is a σ-field on Ω if it satisfies the following conditions:

1. $\emptyset \in \mathcal{F}$;
2. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$;
3. if $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^\infty A_i \in \mathcal{F}$.

- For example, for any Ω, $\mathcal{F} = \{\emptyset, \Omega\}, \mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ and $\mathcal{F} = 2^\Omega$ are all σ-fields.
- For σ-field $\mathcal{F}$, if $A, B \in \mathcal{F}$, then $A \cap B, A \setminus B \in \mathcal{F}$.
  Proof: $A \cap B = (A^c \cup B^c)^c$ and $A \setminus B = A \cap B^c$.
- If $\mathcal{F}_1, \mathcal{F}_2$ are σ-fields, then $\mathcal{F}_1 \cap \mathcal{F}_2$ is also a σ-field. (Try to proof as a homework)
- However, $\mathcal{F}_1 \cup \mathcal{F}_2$ may not be a σ-field. Consider $\mathcal{F}_1 = \{\emptyset, A, A^c, \Omega\}, \mathcal{F}_2 = \{\emptyset, B, B^c, \Omega\}$.
- Question: How $\mathcal{F}$ looks like if $\Omega = \mathbb{R}$ or $\Omega = [a, b]$?

What We Need for Probability

- Suppose that we perform the experiment for N times and denote by $N(A)$ the number of times event A occurs, then we have $P(A) \approx \frac{N(A)}{N}$.
- Clearly $N(\emptyset) = 0$ and $N(\Omega) = N$. Furthermore, if A and B are disjoint events, then $N(A \cup B) = N(A) \cup N(B)$. This suggests that P should be *countably additive*.

Definition: Probability Measure and Probability Space

Let Ω be a sample space and $\mathcal{F} \subseteq 2^\Omega$ be a σ-filed on Ω. A function $P : \mathcal{F} \to [0, 1]$ is said to be a probability measure on $(\Omega, \mathcal{F})$ if it satisfies the following conditions:

1. $P(\emptyset) = 0, P(\Omega) = 1$;
2. for any events $A_1, A_2, \dots \in \mathcal{F}$ such that $\forall i \neq j : A_i \cap A_j = \emptyset$, we have

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Then 3-tuple $(\Omega, \mathcal{F}, P)$ is said to be a **probability space** if $\mathcal{F} \subseteq 2^\Omega$ is a σ-filed on Ω and P is a probability measure on $(\Omega, \mathcal{F})$.

- **Remark (Atom)**
  Sometimes it is not convenient to list all events in $\mathcal{F}$. But since $\mathcal{F}$ is a σ-field, if $A, B \in \mathcal{F}$, then we must have $A \cup B \in \mathcal{F}$. Therefore, it suffices to list A and B. Given $(\Omega, \mathcal{F}, P)$, we say $A \in \mathcal{F}$ is an **atom** if $P(A) > 0$ and $\forall B \subset A : P(B) < P(A) \Rightarrow P(B) = 0$. For example, for $\mathcal{F} = \{\emptyset, \Omega, A, B, C, A \cup B, B \cup C, A \cup C\}$, atoms are A, B and C.

Properties of Probability Space

- For any $A \in \mathcal{F}$, we have $P(A^c) = 1 - P(A)$.
  Proof: (i) $P(A \cup A^c) = P(\Omega) = 1$; (ii) $P(A \cup A^c) = P(A) + P(A^c)$ since $A \cap A^c = \emptyset$.
- If $A_1, A_2 \in \mathcal{F}$ and $A_1 \subseteq A_2$, then $P(A_2) = P(A_1) + P(A_2 \setminus A_1) \geq P(A_1)$.
  Proof: Because A_1 and $A_2 \setminus A_1$ are disjoint.
- For $A_1, A_2 \in \mathcal{F}$, we have $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.
  Proof: $P(A_1 \cup A_2) = P(A_1 \cup (A_2 \setminus A_1)) = P(A_1) + P(A_2 \setminus A_1) = P(A_1) + P(A_2 \setminus (A_1 \cap A_2))$
- For $A_1, A_2 \dots, A_n \in \mathcal{F}$, we have

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

  Proof: Leave as a homework (Hint: by induction)
- For any sequence of events $A_1, A_2, \dots \in \mathcal{F}$, we have

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\lim_{n\to\infty} \bigcup_{k=1}^{n} A_k\right) = \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} A_k\right) \text{; the same for } \cap$$

Conditional Probability

- Suppose that we perform an experiment N times. Then the **conditional probability** for "*if B occurs, then the probability of A is p*" can be counted as $\frac{N(A\cap B)}{N(B)} = \frac{N(A\cap B)/N}{N(B)/N} \approx \frac{P(A\cap B)}{P(B)}$. This motivates the define conditional probability.

Definition: Conditional Probability

If $P(B) > 0$, then the conditional probability that A occurs given that B occurs is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$P(A)$ is the prior probability of A; $P(A \mid B)$ is the posterior probability of A given B.

- **Law of Total Probability & Bayes' Rule**
  Let $A_1 \dot\cup A_2 \dot\cup \ldots \dot\cup A_n = \Omega$ be a partition of Ω and $A_i \in \mathcal{F}$. Then for any $B \in \mathcal{F}$, we have

$$\boxed{P(B)} = P(B \cap \Omega) = P(B \cap (A_1 \cup A_2 \cup \ldots A_n)) = \sum_{i=1}^{n} P(B \cap A_i) = \boxed{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}$$

This also leads to the well-known the Bayes' Rule

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)} \quad \text{or} \quad P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)}$$

Independence

- In general $P(A \mid B)$ changes when B changes unless A does not depend on B. Let us consider the follows equivalence

$$\boxed{P(A \mid B) = P(A \mid B^c)} \Leftrightarrow \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B^c)}{P(B^c)} \Leftrightarrow \frac{P(A \cap B)}{P(B)} = \frac{P(A) - P(A \cap B)}{1 - P(B)} \Leftrightarrow P(A \cap B) = \boxed{P(A)P(B)}$$

Definition: Independence

We say two events $A, B \in \mathcal{F}$ are (statistically) independent if $P(A \cap B) = P(A)P(B)$.

A set of events $A_1, A_2, \ldots, A_n \in \mathcal{F}$ are said to be **mutually independent** if

$$\forall \{k_1, k_2, \ldots, k_l\} \subseteq \{1, 2, \ldots, n\} : P(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_l}) = \prod_{i=1}^{l} P(A_{k_i})$$

- **Remark: Pairwise independence does not imply mutual independence**
  Let $\Omega = \{1, 2, \ldots, 7\}, \mathcal{F} = 2^{\Omega}, P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = \frac{1}{8}, P(\{7\}) = \frac{1}{4}$. Consider $A = \{1, 2, 7\}, B = \{3, 4, 7\}, C = \{5, 6, 7\}$, so $P(A) = P(B) = P(C) = \frac{1}{2}$. However, these three events are

  (1) pairwise independent, since $P(A \cap B) = P(B \cap C) = P(A \cap C) = \frac{1}{4}$;

  (2) not mutually independent, since $P(A \cap B \cap C) = \frac{1}{4} \neq P(A)P(B)P(C) = \frac{1}{8}$.

Examples for Conditional Probability & Independence

- **Example 1**

  There are two kinds of COVID-19 vaccines distributed randomly:

  - 70% are the α-vaccines and 30% are the β-vaccines.
  - The effectiveness of the α-vaccine is 90% and that of the β-vaccine is 50%.

  The experiment is that a person takes a vaccine randomly and waits for the effectiveness.

  Then the sample space is $\Omega = \{C+, C-, U+, U-\}$.

  Let $A = \{C+, U+\}$ be the event that he becomes immune and $B = \{U+, U-\}$ be the event that he takes the β-vaccine. Then we have

  $$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c) = 0.5 \times 0.3 + 0.9 \times 0.7 = 0.78$$

  If the vaccine fails, then the probability that he took the β-vaccine is

  $$P(B \mid A^c) = \frac{P(B \cap A^c)}{P(A^c)} = \frac{0.3 \times 0.5}{1 - 0.78} = 68.2\%$$

  Since $P(A \cap B) = 0.5$ and $P(A) = 0.78$, events A and B are clearly dependent.

- **Example 2**

  Choose a card randomly from a pack of 52 cards. Then we have

  $$P(\mathrm{A}) = \frac{4}{52}, \quad P(\clubsuit) = \frac{1}{4} \quad \text{and} \quad P(\clubsuit\mathrm{A}) = \frac{1}{52}$$

  So we conclude that the suit of the choose card is independent of its rank.