1 Abstract Definition of Probability

Probability is the area of mathematics concerned with describing uncertain or random events. We will develop a mathematical framework that will allow us to quantify uncertainty in a principled way.

1.1 Probability Space

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of three parts: Ω denotes the set of outcomes, $\mathcal{F}$ the events we can assign probability to, and $\mathbb{P}$ the probability of each event. This forms the foundation of probability theory.

The collection of possible outcomes is called the probability space, from which we can associate possible events and their respective probabilities.

Definition 1. A nonempty set Ω is called a **probability space**. The elements of Ω are called outcomes.

Unless Ω is countable, the space of subsets is often too rich to understand. However, if we restrict ourselves to a large, yet natural collection of subsets, we will still retain enough sets to build a rich theory of probability from essential ideas in analysis.

Definition 2. A collection of subsets $\mathscr{F}$ of Ω is called a **σ-algebra** on Ω if

1. $\Omega \in \mathscr{F}$
2. Closed under complements: if $A \in \mathscr{F}$, then $A^c \in \mathscr{F}$
3. Closed under countable unions: if $A_1, A_2, \ldots \in \mathscr{F}$, then $\displaystyle\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}$.

The elements of $\mathcal{F}$ are often called **events**.

Remark 1. When $\Omega = \mathbb{R}^n$, it is typically not possible to take $\mathcal{F}$ to be the power set consisting of all subsets of $\mathbb{R}^n$. Some sets in $\mathbb{R}^n$ are very weird, and need to be excluded. The classical choice of σ-algebra on $\mathbb{R}^n$ is the **Borel σ-algebra**

$$\mathscr{B}(\mathbb{R}^n),$$

which is defined as the smallest σ-algebra that contains all open cubes

$$(a_1, b_1) \times \cdots \times (a_n, b_n) \qquad \text{with } a_i < b_i$$

It is quite difficult to construct sets that are not Borel sets., so $\mathscr{B}(\mathbb{R}^n)$ contains all “reasonable” sets.

The main consequence of restricting our study to σ-algebras is the existence of a way to measure the likelihood of events.

Definition 3. A function $\mathbb{P}$ is a **probability measure** on the σ-algebra $\mathcal{F}$ if

1. $\mathbb{P}(\Omega) = 1$;
2. $\mathbb{P} \geq 0$ for all $A \in \mathcal{F}$;
3. Countable Additivity: if $A_1, A_2, \ldots \in \mathscr{F}$ are disjoint ($A_i \cap A_j = \emptyset$ for all $i \neq j$), then

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

1.2 Properties

From these three properties, we can recover all the natural properties a probability should satisfy.

Proposition 1 (*Properties of a Probability Measure*)

Any probability measure $\mathbb{P}$ satisfies the following

1. $\mathbb{P}(A) \in [0, 1]$,
2. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
3. *Monotonicity: If* $A \subseteq B$, *then* $\mathbb{P}(A) \leq \mathbb{P}(B)$,
4. *Inclusion–Exclusion Principle:* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,
5. *Union Bound:* $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

1.2.1 Discrete Probability Spaces

If Ω is countable, then we can always assign a probability to every outcome. Thus a probability measure is completely determined by the probabilities of each individual outcome.

Corollary 1

Let $\Omega = \{\omega_1, \omega_2, \dots\}$ *and let* $A \subset \Omega$ *be an event. Then*

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\omega_i).$$

Remark 2. On the other hand, if Ω is uncountable, then it is impossible to assign a probability to every outcome, so it is necessary to define probabilities on the set of measurable events $\mathcal{F}$.

A special case of a discrete probability space with finitely many elements $\Omega = \{\omega_1, \dots, \omega_n\}$ is the one with *equally likely outcomes*. In this case, $\mathbb{P}(\omega_1) = \mathbb{P}(\omega_2) = \cdots = \mathbb{P}(\omega_n)$ so the normalization requirement implies that for any index $k \in \{1, \dots, n\}$

$$1 = \mathbb{P}(\Omega) = \sum_{\omega_i \in \Omega} \mathbb{P}(\omega_i) = |\Omega| \, \mathbb{P}(\omega_k) \implies \mathbb{P}(\omega_k) = \frac{1}{|\Omega|}.$$

This is often the most naive definition of a probability space.

Definition 4 (Uniform Probability Measure)**.** Let Ω be finite. If all outcomes are equally likely, then the associated probability measure $\mathbb{P}$ on Ω is called the *uniform probability measure* and

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\omega_i) = \sum_{\omega_i \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}.$$

This means that for uniform probability measures, we can simply count the number of ways an event can happen to compute the probabilities on finite probability spaces. However, we will encounter much richer probability spaces throughout this course, since outcomes might not always be equally likely. Understanding these probabilities will require tools beyond counting.

1.3 Example Problems

Problem 1.1. Suppose two six sided dice are rolled, and the number of dots facing up on each die is recorded.

1. Write down the probability space Ω.
2. Write down, as a set, the event $A =$ "The sum of the dots is 7".
3. Write down, as a set, the event B^c, where $B =$ "The sum of the numbers is at least 4".
4. Write down, as a set, the events $A \cap B^c$ and $A \cup B^c$.

Solution 1.1.

1. The sample space for a pair of dice is the a pair of the outcomes of each die roll
$$\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\} = \{(x, y) : x, y \in \{1, 2, \dots, 6\}\}.$$
2. We can simply write down all the combinations
$$A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}.$$
3. If $B = \{\text{sum is at least } 4\}$ then $B^c = \{\text{sum is at most } 3\}$, so
$$B^c = \{(1, 1), (1, 2), (2, 1)\}.$$
4. Since it is impossible for the sum of dots to be 7 and at most 3 at the same time, $A \cap B^c = \emptyset$. All the possibilities the sum of dots is 7 or at most 3 is
$$A \cup B^c = \{\underbrace{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)}_{A}, \underbrace{(1, 1), (1, 2), (2, 1)}_{B^c}\}.$$

Problem 1.2. For the following experiments, describe a possible probability space Ω.

1. Roll a die.
2. Number of coin-flips until heads occurs.
3. Waiting time in minutes (with infinite precision, e.g., $0.238\overline{445}$ minutes) until a task is complete.

Solution 1.2.

1. There are many ways we can record the outcome of a die such that no elements can occur at the same time $\Omega = \{1, 2, 3, 4, 5, 6\}$ or $\Omega = \{\text{even}, \text{odd}\}$. The choice of the best probability space will depend on the application in mind, but usually the coarsest choice is the most powerful.
2. There is only one natural choice here $\Omega = \{1, 2, 3, \dots\} = \mathbb{N}$.
3. There is only one natural choice here $\Omega = [0, \infty) = \{x \in \mathbb{R} : x \geq 0\}$.

Problem 1.3. Suppose that two fair six sided die are rolled.

1. What is the probability that the dots on each die match?

2. What is the probability that the dots sum to 7?
3. What is the probability that the dots do not sum to 7?
4. What is the probability that the dots match and sum to 7?

Solution 1.3. The probability is uniform over the probability space $\Omega = \{1, \dots, 6\}^2$. All outcomes are equally likely, so for an event A,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{36}.$$

1. The event is $A = \{(1,1),(2,2),\dots,(6,6)\}$ with $|A| = 6$, so $\mathbb{P}(A) = 6/36 = 1/6$.
2. The event is $B = \{(1,6),(2,5),\dots,(6,1)\}$ with $|B| = 6$, so $\mathbb{P}(B) = 6/36 = 1/6$.
3. The event is B^c with $|B^c| = |\Omega| - |B| = 30$ elements, hence $\mathbb{P}(B^c) = 30/36 = 5/6 = 1 - \mathbb{P}(B)$.
4. 7 is an odd number so it is impossible for the dots to match. Therefore, $\mathbb{P}(\emptyset) = 0$.

Problem 1.4. Supposes that $\mathbb{P}(A) = \frac{1}{3}$ and $\mathbb{P}(B^c) = \frac{1}{4}$, can A and B be disjoint?

Solution 1.4. We have

$$\mathbb{P}(B) = 1 - \mathbb{P}(B^c) = \frac{3}{4}.$$

By the inclusion exclusion property, we have

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

Suppose that A and B are disjoint, i.e. $\mathbb{P}(A \cap B) = 0$, then we must have

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) = \frac{1}{3} + \frac{3}{4} = \frac{13}{12} > 1$$

which contradicts the fact that probabilities are between 0 and 1. Thus, A and B cannot be disjoint.

Problem 1.5. Let $\mathcal{F}$ be a σ-algebra. Show that $A, B \in \mathcal{F} \implies A \cap B \in \mathcal{F}, A \cup B \in \mathcal{F}$

Solution 1.5. Since $\mathcal{F}$ is closed under countable unions we have closure under finite unions $A \cup B \in \mathcal{F}$ (by taking $A_3 = A_4 = \cdots = B$).

To see that $A \cap B \in \mathcal{F}$, notice that $A^c, B^c \in \mathcal{F}$ is so $A^c \cup B^c \in \mathcal{F}$ by above. By taking complements we have that $(A^c \cup B^c)^c = A \cap B \in \mathcal{F}$ using De Morgan's identity.

Problem 1.6. Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two σ-algebras on the same probability space. Show that $\mathcal{F}_1 \cap \mathcal{F}_2$ is a σ-algebra.

Solution 1.6. We check the 3 properties.

1. We have that $\Omega \in \mathcal{F}_1$ and $\Omega \in \mathcal{F}_2$, so $\Omega \in \mathcal{F}_1 \cap \mathcal{F}_2$.
2. If $A \in \mathcal{F}_1 \cap \mathcal{F}_2$, then $A^c \in \mathcal{F}_1$ and $A^c \in \mathcal{F}_2$ since $\mathcal{F}$ is closed under complements, so $A^c \in \mathcal{F}_1 \cap \mathcal{F}_2$.
3. Let $A_1, A_2, \dots \in \mathcal{F}_1 \cap \mathcal{F}_2$. Then $\cup_{i=1}^\infty A_i \in \mathcal{F}_1$ and $\cup_{i=1}^\infty A_i \in \mathcal{F}_2$ is closed under countable unions, so $\cup_{i=1}^\infty A_i \in \mathcal{F}_1 \cap \mathcal{F}_2$.

1.4 Proofs of Key Results

Problem 1.7. (Proposition 1) Show the *monotonicity* property of probability,

$$\text{if } A \subseteq B \text{ then } \mathbb{P}(A) \leq \mathbb{P}(B).$$

Solution 1.7. This follows directly from the axioms. If $A \subseteq B$, then

$$B = (A \cap B) \cup (A \setminus B) = A \cup (A \setminus B)$$

and the sets A and $A \setminus B$ are disjoint. Therefore, by countable additivity,

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(A \setminus B) \geq \mathbb{P}(A)$$

since $\mathbb{P}(A \setminus B) \geq 0$ by the non-negativity property.

Problem 1.8. (Proposition 1) Show that the axiomatic defintiion of a probability implies that

$$0 \leq \mathbb{P}(A) \leq 1$$

for any event A.

Solution 1.8. By the monontonicity property, since $A \subseteq \Omega$,

$$\mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$$

by the normalization property. On the other hand, but non-negativity, we have that $\mathbb{P}(A) \geq 0$.

Problem 1.9. (Proposition 1) Show that the axiomatic definition of a probability implies that

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$

for any event A.

Solution 1.9. Notice that $A \cup A^c = \Omega$ and A and A^c are disjoint. From finite additivity, we conclude that

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

Problem 1.10. (Proposition 1) Show that the axiomatic definition of a probability implies that

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

for any event A and B.

Solution 1.10. Notice that A and $(B \setminus A)$ are disjoint events such that $A \cup (B \setminus A) = A \cup B$, so by countable additvity

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

Next, notice that $B \setminus A$ and $A \cap B$ are exclusive and $(A \cap B) \cup (B \setminus A) = A \cap B$,

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

which implies our result.

Problem 1.11. (Proposition 1) Show that the axiomatic definition of a probability implies that

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

for any event A and B.

Solution 1.11. By the inclusion–exclusion property,

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

since $\mathbb{P}(A \cap B) \geq 0$. Notice that this implies that the union bound is sharp and is attained when A and B are disjoint sets, which is implied by countable additivity.