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 Axioms of Probability

The fundamental object is the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which specifies the space of outcomes Ω, the space of events $\mathcal{F}$, and the likelihoods of events $\mathbb{P}$.

Definition 1 (Probability Space)**.** A *sample space* Ω is the *set* of all possible outcomes of a random process. The elements of $\omega \in \Omega$ are called *outcomes* and the subsets $A \subseteq \Omega$ are called *events*. The associated *probability measure* $\mathbb{P}$ encodes the probability of each event occurring.

Definition 2 (Probability Measure)**.** Let $\mathcal{F}$ denote the set of all subsets of Ω that we can assign probabilities to. A probability measure is a function from $\mathcal{F} \to \mathbb{R}_+$ such that

1. Normalization: $\mathbb{P}(\Omega) = 1$
2. Non-Negativity: $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{F}$.
3. Countable Additivity: If $A_1, A_2, ... \in \mathcal{F}$ are disjoint ($A_i \cap A_j = \emptyset$ for $i \neq j$), then
$$\mathbb{P}\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

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: $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.*

1.1.1 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 1. 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 is the one with *equally likely outcomes*. This is often the most naive definition of a probability space. We will encounter much richer probability spaces throughout this course, since outcomes might not always be equally likely.

Definition 3 (Uniform Probability Measure)**.** If all outcomes are equally likely, then the associated probability measure $\mathbb{P}$ is called the *uniform probability measure* and

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

1.2 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 sample 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 sample space Ω.

1. Roll a die.
2. Number of coin-flips until heads occurs.
3. Waiting time in minutes (with infinite precision, e.g., $0.238\overline{45}$ 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 sample 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 sample 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 $|B^c| = |S| - |B| = 30$ 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, $\mathbb{P}(\emptyset) = 0$.

1.3 Proofs of Key Results

Problem 1.4. (Proposition 1) Show the *monotonicity* property of probability,

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

Solution 1.4. This follows directly from the axioms. If $A \subseteq B$, then $B = A \cup A \setminus B$ 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 1.5. (Proposition 1) Show that the axiomatic defintiion of a probability implies that

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

for any event A.

Solution 1.5. Suppose for the sake of contradiction that $\mathbb{P}(A) > 1$ for some event A. By the monontonicity property, since $A \subseteq \Omega$,

$$\mathbb{P}(\Omega) \geq \mathbb{P}(A) > 1$$

which contradicts the fact that $\mathbb{P}(\Omega) = 1$. Therefore, $\mathbb{P}(A) \leq 1$.

Problem 1.6. (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 1.6. 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 1.7. (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 1.7. 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 1.8. (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 1.8. 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.