1 Introduction to Probability

Probability is the area of mathematics concerned with describing uncertain or random events.

1.1 Problematic Definitions of Probability

1. **Classical Definition:** The probabity of an event is

$$\frac{\text{number of ways the event can occur}}{\text{the total number of possible outcomes}}$$

   assuming that all outcomes are equally likely.

   Example: The probability of flipping a heads on a coin is $\frac{1}{2}$ because flipping a H is one out of only two possible outcomes H and T.

   Downfalls: This definition is not easily extendable to events that are not equally likely. If we could, then everything has a 50% chance of happening: it either happens or it doesn't.

2. **Relative Frequency:** The probability of an event is the proportion of times the event occurs in a very long series of experiments.

   Example: If everyone in the class flips a coin, then H will show up roughly $\frac{1}{2}$ of the time.

   Downfalls: It is often impossible to repeat certain experiments infinitely often.

3. **Subjective:** The probability of an event is a measure of how sure the person making the statement is that the event will happen.

   Example: There is no reason for one H or T to be more likely because the coin is symmetric, so the probability of flipping a H is $\frac{1}{2}$.

   Downfalls: This definition is not objective. Everyone can have wildly different opinions on correct probabilities for more complicated events.

1.2 Mathematical Probability Model

To overcome the shortcomings of the previous definition, we instead treat probability as a mathematical system defined by a set of axioms. We will treat probabilities as ways to measure objects in sets.

Definition 1 (Review of Set Notation)**.** Let A, B be sets.

1. *Element*: $x \in A$ if the outcome x is in the event A.
2. *Empty set*: The empty set is denoted $\emptyset$
3. *Complement*: $A^c = \{x \mid x \in S,\ \ x \notin A\}$ (often A^c is also denoted by $\overline{A}$)
4. *Cardinality:* $|A|$ is the number of elements in the set A
5. *Union*: $A \cup B = \{x \mid x \in A \ \ or \ \ x \in B\}$
6. *Intersection*: $A \cap B = \{x \mid x \in A \ \ and \ \ x \in B\}$
7. *Set Difference*: $A \setminus B = A \cap B^c = \{x \mid x \in A \text{ and } x \notin B\}$

8. *Cartesian Product:* $A \times B = \{(x,y) \mid x \in A \text{ and } y \in B\}$. Recall that in contrast to the set notation $\{x,y\}$, the notation (x,y) refers to an ordered set, so $(x,y) \neq (y,x)$ if $x \neq y$ while $\{x,y\} = \{y,x\}$ even when $x \neq y$.

9. *Disjoint*: Two events A and B are said to be disjoint if $A \cap B = \emptyset$.

Definition 2. A *sample space* S is a *set* of distinct outcomes of an experiment with the property that in a single trial of the experiment only one of these outcomes occurs.

Definition 3. A set A is an *event* if $A \subseteq S$, for which we want to assign probabilities to.

Definition 4 (Axioms of Probability)**.** Let $\mathcal{S}$ denote the set of all events on a given sample space S. A *probability* defined on $\mathcal{S}$ is a function

$$\mathbb{P} : \mathcal{S} \to \mathbb{R},$$

that satisfies the following three conditions:

1. Non-Negativity: If A is an event, then $\mathbb{P}(A) \geq 0$.

2. Normalization: $\mathbb{P}(S) = 1$

3. Countable Additivity: If $A_1, A_2, ...$ is a sequence of disjoint events, that is, $\mathbb{P}(A_i \cap A_j) = \emptyset$ for $i \neq j$, then

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

Remark 1. The course textbooks uses a stronger form of the non-negativity assumption: $0 \leq \mathbb{P}(A) \leq 1$. However, the fact that all probabilities are less than 1 is implied by the definition stated here.

Definition 5. *Odds* are another way to describe probabilities, which encodes the ratio of how likely a probability is relative to its complement

1. Odds **in favour** of an event A occurring is

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

2. Odds **against** an event A is

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

1.3 Discrete Probability Spaces

Definition 6. A sample space S is *discrete* if S is countably infinite. We say that a sample spaces S is *finite* if $|S|$ is finite. Of course, this means that all finite sample spaces are discrete, but not all discrete sample spaces are finite.

Definition 7. Let S be discrete and $A \subset S$ an event. If A contains only one point, we call it a *simple event*, otherwise it is called a *compound event*.

Definition 8. Let $S = \{a_1, a_2, \dots\}$ discrete and $A \subset S$ an event. Then

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

Definition 9 (Discrete Probability Measure)**.** Let $S = \{a_1, a_2, \dots\}$ be discrete. Assign numbers to each of the individual outcomes

$$\mathbb{P}(\{a_i\}) = \mathbb{P}(a_i) = p_i, \quad i = 1, 2, \dots$$

so that

1. $p_i \geq 0, \quad i = 1, 2, \dots$
2. $\sum_i p_i = 1$.

For any event $A \subseteq S$, its probability is given by

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

We then call the set of probabilities (that satisfy condition 1 and 2) $\{\mathbb{P}(a_i) \mid i = 1, 2, \dots\}$ a *probability distribution.*

Definition 10. We say a finite sample space $S = \{a_1, \dots, a_n\}$ has *equally likely* outcomes if the probability of every individual outcome in S is the same. That is,

$$\mathbb{P}(a_i) = p \text{ for all } i \leq n \implies p = \frac{1}{|S|}$$

since $\sum_{i=1}^n \mathbb{P}(a_i) = np = 1$ and $|S| = n$. We also say that the probability $\mathbb{P}$ is *uniform* on S. If the outcomes are equally likely, it follows that

$$\mathbb{P}(A) = \frac{|A|}{|S|}$$

which coincides with the classical definition of probability.

1.4 Example Problems

1.4.1 Basic Definitions

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 S.
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

$$S = \{1, \dots, 6\} \times \{1, \dots, 6\} = \{(x, y) : x, y \in \{1, 2, \dots, 6\}\}$$

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. It follows that $A \cap B^c = \emptyset$ and

$$A \cup B^c = \{1,6),(2,5),(3,4),(4,3),(5,2),(6,1),(1,1),(1,2),(2,1)\}$$

Problem 1.2. For the following experiments, describe a possible sample space S.

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 $S = \{1,2,3,4,5,6\}$ or $S = \{\text{even}, \text{odd}\}$. 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 $S = \{1,2,3,\dots\} = \mathbb{N}$
3. There is only one natural choice here $S = [0,\infty) = \{x \in \mathbb{R} : x \geq 0\}$

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 $S = \{1,\dots,6\}^2$. This is an equally likely sample space, hence, for an event A,

$$P(A) = |A|/|S| = |A|/36.$$

1. The event is $A = \{(1,1),(2,2),\dots,(6,6)\}$ with $|A| = 6$, so $P(A) = 6/36 = 1/6$.
2. The event is $B = \{(1,6),(2,5),\dots,(6,1)\}$ with $|B| = 6$, so $P(B) = 6/36 = 1/6$.
3. The event is B^c with $|B^c| = |S| - |B| = 30$ elements, hence $P(B^c) = 30/36 = 5/6 = 1 - P(B)$.
4. The event is $A \cap B = \emptyset$, hence, $P(A \cap B) = P(\emptyset) = 0$.

Problem 1.4. A fair six-sided die is rolled once. What are:

1. the odds in favour of rolling a 6?
2. the odds against rolling a 6?

3. the odds in favour of rolling an even number?

Solution 1.4.

1. Let $A = \{\text{rolling a } 6\}$. We have $\mathbb{P}(A) = \frac{1}{6}$ and $1 - \mathbb{P}(A) = \frac{5}{6}$ so the odds in favor of A are $1/5$. which is commonly written as $1 : 5$.
2. Let $A = \{\text{rolling a } 6\}$. We have $\mathbb{P}(A) = \frac{1}{6}$ and $1 - \mathbb{P}(A) = \frac{5}{6}$ so the odds against A if $5/1$ which is commonly written as $5 : 1$.
3. Let $A = \{\text{rolling an even number}\}$. We have $\mathbb{P}(A) = \frac{1}{2}$ and $1 - \mathbb{P}(A) = \frac{1}{2}$ so the odds in favor of A are $1/1$ which is commonly written as $1 : 1$.

1.4.2 Properties of Probability Measures

Problem 1.5. $(*)$ Show the *monotonicity* property of probability,

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

Solution 1.5. 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 1.6. $(*)$ Show that the axiomatic defintiion of a probability implies that

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

for any event A.

Solution 1.6. Suppose for the sake of contradiction that $\mathbb{P}(A) > 1$ for some event A. By the monontonicity property, since $A \subseteq S$,

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

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

Problem 1.7. $(*)$ Show that the axiomatic defintiion of a probability implies that

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

for any event A.

Solution 1.7. Notice that $A \cup A^c = S$ and A and A^c are disjoint. From finite additivity, we conclude that

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

Problem 1.8. $(*)$ Let $S = \{a_1, a_2, \dots, a_n\}$ be a finite sample space. We define the function $A \subset S$ by

$$f(A) = \frac{|A|}{|S|}.$$

Show that the function f defines a discrete probability measure $\mathbb{P}$ on S. Furthermore, show that $\mathbb{P}$ is uniform on S.

Solution 1.8. It suffices to check that the function f satisfies the 3 axiomatic conditions of a probability measure.

1. Non-negative: Since the cardinality is non-negative $f(A) = \frac{|A|}{|S|} \geq 0$.

2. Normalization: We have $f(S) = \frac{|S|}{|S|} = 1$

3. Countable Additivity: Since our sample space is finite, it suffices to show finite addivity. If $A_1, \dots, A_k$ are disjoint events, then the definition of the cardinatliy of the set satisfies

$$|A_1 \cup \cdots \cup A_k| = |A_1| + \cdots + |A_k|.$$

   Therefore,

$$\mathbb{P}\left(\bigcup_{i=1}^{k} A_i\right) = \frac{|A_1 \cup \cdots A_k|}{|S|} = \sum_{i=1}^{k} \frac{|A_i|}{|S|} = \sum_{i=1}^{k} \mathbb{P}(A_i)$$

Therefore, f defines a probability measure $\mathbb{P}$ on S. Furthermore, for any elements a_i, we have

$$\mathbb{P}(a_i) = \frac{|\{a_i\}|}{|S|} = \frac{1}{|S|}$$

so the probability is uniform. This is a useful example to keep in mind because many counting techniques in the next section can be interpreted as fundamental operations for probabilities.