Event and Probability Axioms

1 Introduction

Many phenomena in life are unpredictable. This is what makes games (and perhaps life) appealing. The unpredictability leads to words and concepts such as randomness, chance, and probability. For example, we say the chance of rain tomorrow is 30% or the probability of Miami Heat winning the playoffs is small.

Probability theory is way to formalize and assign quantitive meanings to these words and concepts. The theory should lead to results that are in agreement to our practical experimentation. This week, we study the foundation of probability theory.

2 Events

An **experiment** or a *trial* is a sequence of actions or happenings whose **outcome** is not fully predictable in advance, for example, tossing a coin or rolling a die. An **event** is a set of possible outcomes, for example, heads showing or the die landing on an odd number. We say that an **event occurs** if the outcome of the experiment is in the event. The collection of all possible outcomes is called the **sample space** and it is typically denoted by Ω.

Let us consider the experiment of tossing a coin. There are two possible outcomes: heads (denoted by H) and tails (denoted by T). We write this as

$$\Omega = \{H, T\}.$$

Each subset of Ω *is an event* since each subset is a set of possible outcomes. Thus the events are: $\{\}, \{H\}, \{T\}, \{H, T\}$.

As another example, consider the experiment of rolling a die. There are six outcomes depending on which number is showing. So

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Then

- the event that 1 is showing can be denoted by $A = \{1\}$.
- an odd number is shown can be denoted $B = \{1, 3, 5\}$.
- a number larger than 2 is showing can be denoted by $C = \{3, 4, 5, 6\}$
- odd or larger than 2 $\rightarrow B \cup C = \{1, 3, 4, 5, 6\}$
- odd and larger than 2 $\rightarrow B \cap C = \{3, 5\}$

We already know that events are subsets of Ω. But do we need to include all subsets of Ω? Suppose we only care about the outcome being odd or even. If we only have knowledge of the the outcome being even or odd, then we have information about the following events:

- outcome is odd, $A = \{1, 3, 5\}$.

- outcome is even, $B = \{2, 4, 6\}$.
- outcome is in the empty set, $\varnothing$ (clearly, never happens)
- outcome is in the sample space, Ω (clearly always happens)

Interesting subsets of Ω are included in the set of events. In addition to these, there are other subsets that we include, as explained below.

Suppose that I am interested to know wether the outcome of an experiment is in an event A or not. Then, I am also interested to know wether the outcome is in A^c or not since one statement follows from the other. So if A is an event, then A^c is also an event.

Furthermore, knowledge regarding the occurrence of two events A and B implies knowledge of occurrence of $A \cup B$, so we can reasonably concern ourselves with $A \cup B$. A similar argument holds for $A \cap B$.

Finally, we always know that the empty set never occurs and Ω always occur.

Let the set of events be denoted by $\mathcal{F}$. The above discussion gives intuitive reasoning for the following requirements for $\mathcal{F}$.

Event axioms: The **set of events**, $\mathcal{F}$, must satisfy:

1. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
2. If $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$. More generally, if $A_1, A_2, \cdots$ are in $\mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i$ is also in $\mathcal{F}$.
3. $\Omega \in \mathcal{F}$.

Such a collection of subsets is called a $\sigma-$**algebra.** Clearly, many choices are available for a $\sigma-$algebra. We choose the one that fits our purpose, i.e., the one that contains the events in which we are interested in.

Exercise 1. Show that $\varnothing \in \mathcal{F}$ and if $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$.

If we are interested to know wether the outcome of rolling a die is even or odd, it suffices to have

$$\mathcal{F} = \{\{1, 3, 5\}, \{2, 4, 6\}, \varnothing, \Omega\}.$$

The smallest algebra is $\mathcal{F} = \{\varnothing, \Omega\}$.

3 Probability

We have now defined the sample space and events. But a main component is missing: *probability.* To each event $A \in \mathcal{F}$, we assign a probability $P(A)$. Probability can be interpreted in the following ways:

1. Probability of A, $P(A)$, is a quantity that corresponds to our belief of the chance of occurrence of A. The higher the chance, the larger the probability.
2. Suppose that an experiment is repeated a large number N of times and let $N(A)$ denote the number of times that A occurs. Scientific experience indicates that most of the time, as N grows larger, the ratio $N(A)/N$ settles down. We take the resulting value as the probability of A.

The above interpretations are not contradictory. The second interpretation is more precise than the first but still has its shortcomings. There are other interpretations as well. We use the second interpretation to motivate the **axioms of probability** but after we define our axioms, none of the above interpretations has any role in rigorous arguments.

First, since $N(A) \geq 0$ for any A, we expect $P(A) \geq 0$ as well. Also it is clear that $N(\Omega) = N$, and if A and B are disjoint, then $N(A \cup B) = N(A) + N(B)$. Translating these relations to probabilities we obtain

$$P(A \cup B) = P(A) + P(B), \quad \text{if } A \cap B = \varnothing$$
$$P(\Omega) = 1.$$

These relations are sufficient for determining the desired properties that we want in a probability measure.

Probability axioms: A *probability measure* P on $(\Omega, \mathcal{F})$ is a function with domain $\mathcal{F}$ that satisfies:

1. $P(A) \geq 0$ for any event A
2. For $A, B \in \mathcal{F}$, if $A \cap B = \varnothing$, then $P(A \cup B) = P(A) + P(B)$. More generally, if $A_1, A_2, \cdots$ are disjoint, then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.
3. $P(\Omega) = 1$.

Note that the above axioms only determine the required properties of P and leave a lot of freedom to choose the numerical values. The numerical values should be determined from the description of the experiment. For example, given a *fair* coin, we let $P(\{H\}) = P(\{T\}) = 1/2$.

Exercise 2. Show that

- For any $A \in \mathcal{F}$, $P(A^c) = 1 - P(A)$
- For any $A \in \mathcal{F}$, $P(A) \leq 1$
- $P(\varnothing) = 0$
- For $A, B \in \mathcal{F}$, if $A \subset B$, then $P(A) \leq P(B)$
- For $A, B \in \mathcal{F}$, $P(A \cup B) = P(A) + P(B) - P(A \cup B)$
- Find an expression for $P(A \cup B \cup C)$

Probability with equally likely outcomes

In some experiments, all outcomes are equally likely; we do not have any reason to assume one outcome is more likely than another. In such experiments, for all $\omega \in \Omega$, we assign

$$P(\{\omega\}) = \frac{1}{|\Omega|}.$$

Now consider an event E. Each element of E has probability $1/|\Omega|$. So E has probability

$$P(E) = |E| \times \frac{1}{|\Omega|} = \frac{|E|}{|\Omega|}.$$

Example 3. A fair coin is tossed twice. What is the probability that we see one heads and one tails?

The sample space is $\Omega = \{(H,H), (H,T), (T,H), (T,T)\}$. Let A denote the event of observing a heads and a tails: $A = \{(H,T), (T,H)\}$.

Since the coin is fair, all outcomes are equally likely; we do not have any reason to assume, say, (H,T) is more likely than (T,T). Thus,

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2}{4}.$$