Probability spaces

Probability is a quantity that measures the degree of certainty towards a future event. Probability is a value between 0 and 1 where a probability of 1 means that we are certain a given event *will* occur and a probability of 0 means we are certain it *will not* occur. That is, the closer the probability is towards 1, the more certain we are that the event will occur.

Probability is mathematically modeled by the concept of a **probability space**. A probability space is a measure space $(\Omega, \mathcal{F}, P)$ where $P(\Omega) := 1$. The set Ω is called the **sample space**, the σ-algebra $\mathcal{F}$ is called the **collection of events**, and the measure P is called the **probability measure** (Definition 1).

Modeling real world phenomena with probability theory requires us to explicitly define a specific experiment on a specific system. Each such possible state or configuration of the system after the experiment is performed constitutes an **outcome** (Definition 1). We then group together these outcomes into sets, called **events**, and then assign each event a probability via the probability measure. That is, given an event $A \subset \Omega$, the quantity $P(A)$ gives us the probability that we will observe an outcome in the event A.

Definition 1 *A* ***probability space*** *is a measure space* $(\Omega, \mathcal{F}, P)$ *where* $P(\Omega) := 1$.

The set Ω *is called the* ***sample space***.

Each element $\omega \in \Omega$ *is called an* ***outcome***.

The σ*-algebra* $\mathcal{F}$ *is called the* ***collection of events***.

Each element $A \in \mathcal{F}$ *is called an* ***event***.

The measure P *is called a* ***probability measure***.

Intuition

Why assign probability to events rather than to outcomes?

One detail in the measure-theoretic model of probability that I found to be confusing is that probability is defined for *events* rather than for *outcomes*. Why is this? Why can't we simply define probability over outcomes? If Ω is countable, then yes, we can simply forget $\mathcal{F}$ and assign probabilities to outcomes. That is, we can simply define P to be a function on outcomes rather than on sets of outcomes and make sure that P is countably additive, positive, and sums to 1 over all elements in Ω. This would be akin to setting

$$\mathcal{F} := \mathcal{P}(\Omega)$$

where $\mathcal{P}$ constructs the powerset.

However, problems start to creep in when Ω is uncountable. For example, take $\Omega := [0, 1]$. In this situation, let's say we want to model the situation in which a number may found anywhere in this region with uniform probability. That is, we have equal certainty that the number can be anywhere in this interval. Naively, we might attempt to assign an equal probability p for each $\omega \in \Omega$ such that $\sum_{\omega \in \Omega} p = 1$. However, clearly this is not possible. If $p > 0$, then the sum of the probabilities of all the outcomes would be ∞ and if $p = 0$, then then sum would be zero. By only allowing the assignment of probabilities to *subsets* of Ω, rather than to individual elements in Ω, we circumvent this issue. In the aforementioned example, we can define probabilities for each subinterval of $[0, 1]$ to be proportional to that subinterval. That is, the probability of finding the number in $[0, 0.5]$ is simply 0.5 because the length of this subinterval is half of the entire interval. In this situation, the set of all subintervals constitutes the σ-algebra.

Why model probability using a measure?

A measure space encapsulates many of the inherent characteristics that are natural for a definition of "uncertainty." Several of the features of a measure space are a natural fit for modeling uncertainty:

- A σ-algebra gives us a way to ensure that if we are discussing our certainty towards an event occurring, we can also discuss our certainty towards the event *not occurring*. That is, the complement of any element of the σ-algebra is also in the σ-algebra.

- The σ-algebra ensures that we can discuss the probability of unions and intersections of events.

- The fact that $P(\emptyset) := 0$ expresses the impossibility that no outcome in the sample space occurs. That is, we are certain to see *some configuration* of the system after the experiment.

- The fact that $P(\Omega) := 1$ expresses the fact that *some* outcome of the sample space will certainly occur. Again, this expresses the fact that we are certain to see *some configuration* of the system after the experiment.

- Intuitively uncertainty is a measurable quantity. If we have two disjoint events, then the probability of either of these events occurring should be the sum of the probabilities of the individual events. By using a measure space to model uncertainty, we ensure that given $A, B \subset \Omega$ where $A \cap B = \emptyset$, it follows that

$$P(A \cup B) = P(A) + P(B)$$

Properties

Probability of the compliment of an event

The probability of an event occurring is 1 minus the probability of the event not occurring (Theorem 1).

Theorem 1 *Given a probability space* $(\Omega, \mathcal{F}, P)$ *where* $A \in \mathcal{F}$*,*

$$P(A) = 1 - P(A^c)$$

Proof:

$$\begin{aligned}
&P(\Omega) = 1 \\
&\Omega = A \cup A^c \\
&P(A \cup A^c) = 1 \\
&P(A \cup A^c) = P(A) + P(A^c) && \text{because } A \text{ and } A^c \text{ are disjoint events} \\
&P(A) + P(A^c) = 1 \\
&P(A) = 1 - P(A^c)
\end{aligned}$$

$\square$

Implication

Given two events A and B where $A \implies B$, the probability of A will be smaller than that of B. The " $\implies$ " operator signifies that whenever A occurs, B *must* have occurred also. That is, $A \subseteq B$. Theorem 2 shows that

$$A \subseteq B \implies P(A) \leq P(B)$$

Theorem 2 *Given a probability space* $(\Omega, \mathcal{F}, P)$ *where* $A, B \in \mathcal{F}$*,*

$$A \subseteq B \implies P(A) \leq P(B)$$

Proof:

$$
\begin{aligned}
&B = A \cup (B \setminus A) && \text{because } B \text{ contains } A \\
&P(B) = P(A) + P(B \setminus A) && \text{because } A \text{ and } B \setminus A \text{ are disjoint sets} \\
&P(B) \geq P(A) && \text{Since } P(X) > 0 \text{ for any } X \in \mathcal{F}
\end{aligned}
$$

□

The probability of an event subtracted from another event

Theorem 3 shows how one can compute the probability of an event obtained by subtracting an event B from an event A.

Theorem 3 *Given a probability space* $(\Omega, \mathcal{F}, P)$ *where* $A, B \in \mathcal{F}$,

$$P(A \setminus B) = P(A) - P(A \cap B)$$

Proof:

$$
\begin{aligned}
&(A \cap B) = (A \setminus B) \cup (A \cap B) && \\
&P(A) = P(A \setminus B) + P(A \cap B) && \text{by definition of a measure} \\
&P(A \setminus B) = P(A) - P(A \cap B) &&
\end{aligned}
$$

□