Introduction to Uncertainty Quantification
Module 2.1: Elements of Set Theory & Probability

1 Elements of Set Theory

To introduce concepts in Probability Theory, it's important that we first define some basic terms in Set Theory. These will serve as building blocks upon which we will establish the Axioms of Probability, and from there begin to explore more advanced concepts.

1.1 Some Definitions

Let's begin by introducing the terms that we'll need throughout this module. For each definition, where applicable, we also introduce the symbol or notation used for each item.

Experiment: Any well-defined action.

Here, when we refer to an experiment this does *not* necessarily refer to a controlled laboratory setting. An experiment, by this broad definition, can be any clearly defined action that has an observed outcome. This could be as simple as a roll of a die or as complex as measuring wind velocity in a hurricane, counting patients as they arrive in an emergency room, or monitoring bacteria levels in a reservoir.

Outcome: (O) Each possible result from the experiment.

Any well-defined experiment will always have observable outcomes. In the roll of a die, there are six possible outcomes. In other experiments, there may be a much larger and potentially infinite number of possible outcomes.

Sample Space: (Ω) The set of all possible outcomes

Here, we will define three different types of samples spaces. These will be very important as we move into our discussion of probability.

1. **Discrete & Finite**: A sample space that has a finite number of outcomes.
   The basic examples of *discrete and finite* sample spaces are a coin flip where $\Omega = \{H, T\}$ and the roll of a standard die where $\Omega = \{1, 2, 3, 4, 5, 6\}$. Many other examples exist, for example the draw of a card from a standard deck, or the draw of 10 cards from a standard deck although the sample space becomes combinatorially large.

2. **Discrete & Infinite**: A sample space that has an infinite, but denumerable (countable) number of outcomes.
   This can be illustrated by the following experiment. Consider that we repeatedly flip coins until we see heads. The sample space, in this case, has an infinite number of outcomes that begin as follows.
   $$\begin{aligned} O_1 &= H \\ O_2 &= TH \\ O_3 &= TTH \\ O_4 &= TTTH \\ &\dots \end{aligned}$$
   In other words, it is possible (though very unlikely) that we can continually roll tails an infinite number of consecutive times.

3. **Continuous**: A sample space that has an infinite and non-denumerable (uncountable) set of outcomes.
   Continuous sample spaces are very common and we will primarily deal with continuous spaces. Examples include any space that occupies the real number line, or some portion of it. For example, any real number in [0, 1] ($\Omega = \{[0, 1]\}$) has an infinite number of outcomes in the interval.

Event: (E) A collection of outcomes. A subset of the sample space Ω.

An event can be any collection of the outcomes in the sample space. Conceptually, we can illustrate events in a *Venn Diagram* as shown in Figure 1 below. As an example, consider that we flip two coins. Some possible events in this sample space include $[H, T]$ and $[H, H]$.

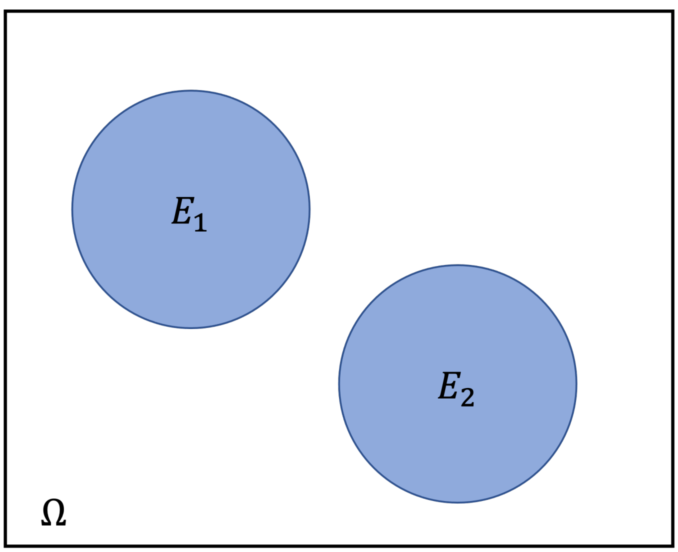

Figure 1: Venn Diagram showing the sample space Ω and two potential Events E_1 and E_2.

Union: ($\cup$) Consider two events E_1 and E_2. The *union* of E_1 and E_2, $E_1 \cup E_2$ is the event that either E_1 *or* E_2 occurs.

The union of two events is illustrated in the Venn Diagram in Figure 2 below.

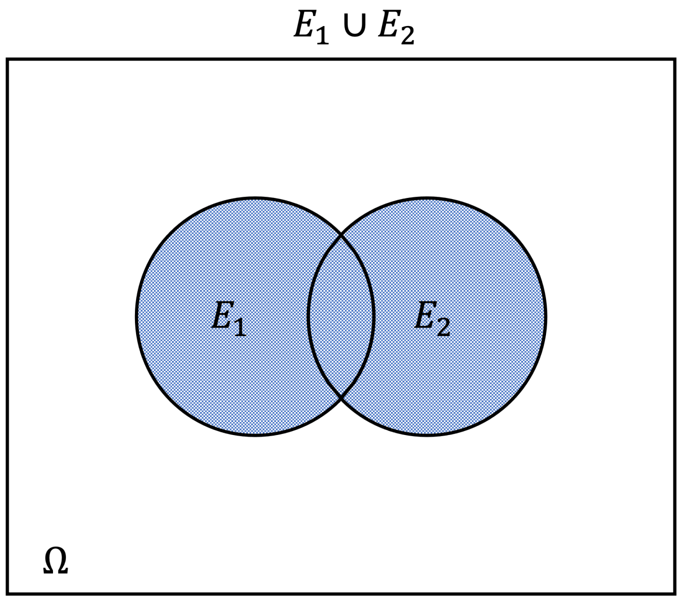

Figure 2: Venn Diagram showing the sample space Ω and the union of two Events E_1 and E_2 as the shaded region.

Intersection: ($\cap$). Consider two events E_1 and E_2. The *intersection* E_1 and E_2, $E_1 \cap E_2$ is the event both E_1 *and* E_2 occur.

The intersection of two events is illustrated in the Venn Diagram in Figure 3 below.

Empty Set / Null Set: ($\emptyset$) The event that contains no outcome.

Initially this may be a difficult concept to grasp. How can we have an event in which there is no outcome? We'll see an example of this in the next definition, and will further see that this concept plays an important role in Set Theory and Probability Theory as we progress. At this stage, suffice it to say that there exists such an event with no outcome.

Mutually Exclusive: Two events E_1 and E_2 are called *mutually exclusive* if $E_1 \cap E_2 = \emptyset$.

As promised, we now see that the empty set can indeed arise. This is illustrated nicely in Figure 1 where we can clearly see that the two events E_1 and E_2 do not intersect. Hence, their intersection is empty. They are therefore referred to as mutually exclusive events. A simple example of mutually exclusive events arises from a single flip of a coin. The events {H} and {T} are mutually exclusive because they cannot both happen in a single flip.

Complement: The complement of an event E_1, denoted E_1^* is the event that includes all possible outcomes that are not in E_1. That is

$$\begin{aligned} E_1 \cup E_1^* &= \Omega \\ E_1 \cap E_1^* &= \emptyset \end{aligned} \tag{1}$$

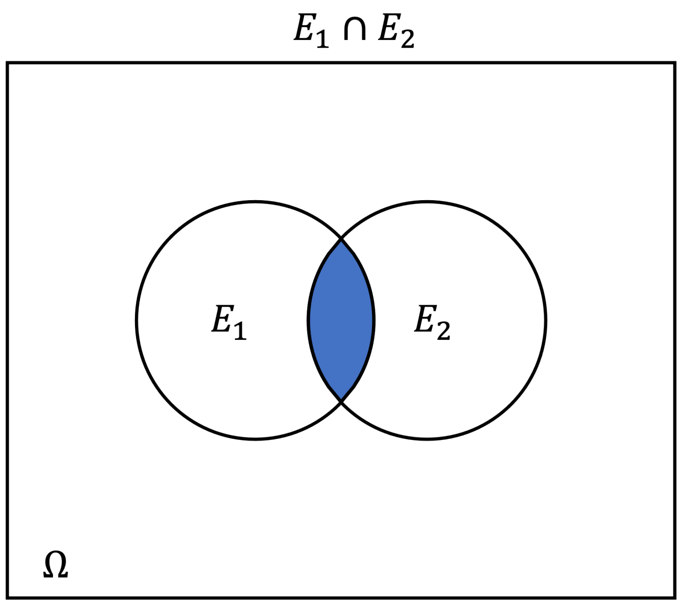

Figure 3: Venn Diagram showing the sample space Ω and the intersection of two Events E_1 and E_2 as the shaded region.

The complement of an event E_1 is illustrated in Figure 4. Again, a simple example of the complement can be seen from the flip of a single coin. The complement of the event {H} is the event {T}.

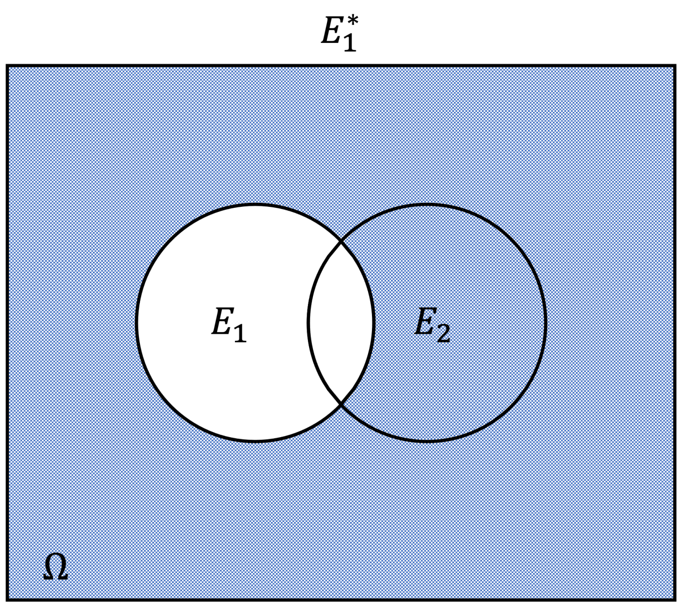

Figure 4: Venn Diagram showing the sample space Ω and the complement of Event E_1 as the shaded region.

1.2 Commutative, Associative, and Distributive Laws

Commutative Laws: The operations of both union and intersection are *commutative*. That is, the result of either operation is does not depend on the order in which the operation is applied. Consider two sets, E_1 and E_2, the *Commutative Laws* of Union and Intersection respectively state that:

$$\begin{aligned} E_1 \cup E_2 &= E_2 \cup E_1 \\ E_1 \cap E_2 &= E_2 \cap E_1 \end{aligned} \tag{2}$$

Associative Laws: The operations of both union and intersection are also *associative*. The *Associative Laws* of Union and Intersection state that the grouping of two or more sets under a given operation does not affect the next set of grouping. Given three events, E_1, E_2, and E_3, this is given by the following relations.

$$\begin{aligned} E_1 \cap (E_2 \cap E_3) &= (E_1 \cap E_2) \cap E_3 \\ E_1 \cup (E_2 \cup E_3) &= (E_1 \cup E_2) \cup E_3 \end{aligned} \tag{3}$$

Distributive Laws: There are two *Distributive Laws* for sets.

First Law: The first law states that taking the union of a set with the intersection of two other sets is equal to taking the union of the first set with each of the other sets separately, and taking the intersection of the results. Given three events, E_1, E_2, and E_3, this is given by the following relation.

$$E_1 \cup (E_2 \cap E_3) = (E_1 \cup E_2) \cap (E_1 \cup E_3) \tag{4}$$

In other words, we can *distribute* the intersection operation to each of the events of the union of two events.

Second Law: The second law states that taking the intersection of a set with the union of two other sets is equal to taking the intersection of the first set with each of the other sets separately, and taking the union of the results. Given three events, E_1, E_2, and E_3, this is given by the following relation.

$$E_1 \cap (E_2 \cup E_3) = (E_1 \cap E_2) \cup (E_1 \cap E_3) \tag{5}$$

In other words, we can *distribute* the union operation to each of the events of the intersection of two events.

In the exercises, you will be asked to illustrate all of these concepts with Venn Diagrams. To assist, the first distributive law is illustrated in Figure 5 below.

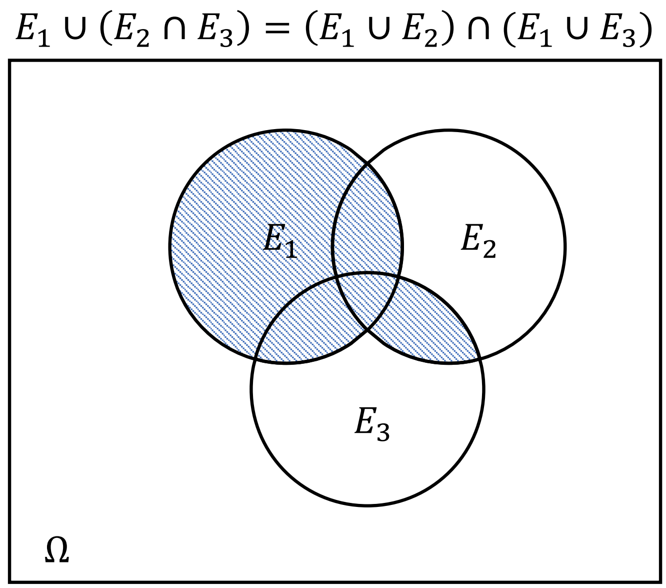

Figure 5: Distributive Law

2 Elements of Probability

In this section, we will introduce the basic elements of probability from the perspective of Set Theory. We begin by defining the abstract space in which we operate when dealing with probabilities, known as the Probability Space.

2.1 Probability Space

A *Probability Space* is composed of three components. For this reason, it is often referred to as a "probability triple" denoted by $(\Omega, \mathcal{F}, P)$. Here, we will introduce each of these components in turn.

Sample Space: (Ω) The set of all possible outcomes.

We have already introduced the *Sample Space*. In probability theory, the sample space is identical to the sample space in set theory. See Section 1.1 above to review the definition and the three types of sample spaces.

One additional note, where applicable we will denote the elements of the sample space by ω.

Event Space: $(\mathcal{F})$ A collection of subsets (or events) of the sample space Ω having the following properties:

1. The empty set is a member of the event space:

$$\emptyset \in \mathcal{F} \tag{6}$$

This property can be equivalently state as follows. The sample space is an element of the event space:

$$\Omega \in \mathcal{F} \tag{7}$$

2. If an event E is a member of the event space, then it's complement is also a member of the event space. This is referred to as being "closed under complements" and can be expressed as:

$$E \in \mathcal{F} \Longrightarrow E^* \in \mathcal{F} \tag{8}$$

3. If a collection of n events E_i for $i = 1, 2, \ldots, n$ are all members of the event space, then the union of those events is a member of the event space. This is referred to as begin "closed under unions" and can be expressed as:

$$E_i \in \mathcal{F} \text{ for } i = 1, 2, \ldots, n \Longrightarrow \cup_{i=1}^{n} E_i \in \mathcal{F} \tag{9}$$

The event space is also often referred to as the σ-field or the σ-algebra.

Probability Measure: (P) A function that maps each event in the event space to a probability, which is a number in [0, 1]. That is:

$$P : \mathcal{F} \rightarrow [0, 1] \tag{10}$$

The *probability measure* must further satisfy the *Axioms of Probability* discussed next.

2.2 Axioms of Probability

The three axioms of probability define the probability measure P and its important properties. These are described next.

1. **Non-negativity:** For any event, $E \in \mathcal{F}$, the probability of E is a non-negative real number. That is:

$$P(E) \geq 0 \quad \forall E \in \mathcal{F} \tag{11}$$

where $P(E)$ is the probability of event E.

2. **Unit Measure:** The probability that at least one event in the sample space will occur equals 1. That is:

$$P(\Omega) = 1 \tag{12}$$

3. **σ-Additivity:** For any set of *mutually exclusive* events E_i, $i = 1, 2, \ldots, n \in \mathcal{F}$, the probabilities are additive. That is:

$$P\left(\cup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) \tag{13}$$

Note that Axiom 3 applies to *mutually exclusive* events. If two events are not mutually exclusive, then

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) \tag{14}$$

Here, we need to ensure that we don't count the intersection of these events twice.

2.3 Resulting Properties

The *Axioms of Probability* result in the following important properties.

1. Probabilities are monotonically non-decreasing.

This means that, for any event $E_1 \in \mathcal{F}$ that is a subset of another event $E_2 \in \mathcal{F}$, i.e. $E_1 \subseteq E_2$, then

$$P(E_1) \leq P(E_2) \tag{15}$$

2. Probabilities are bounded on [0, 1]. That is, for any event $E \in \mathcal{F}$:

$$0 \leq P(E) \leq 1 \tag{16}$$

3. The probability of the empty set, $\emptyset$, is equal to zero. That is:

$$P(\emptyset) = 0 \tag{17}$$

4. If E and E^* are complementary sets (i.e. $E \cup E^* = \Omega$), then

$$P(E) = 1 - P(E^*) \tag{18}$$

3 Conditional Probability

In this section, we will develop the concept of *conditional probability*. Conditional probability relates the probabilities of two events that may, or many not, be dependent on one another. The conditional probability defines the probability of an event E_1, given that some other event E_2 has occurred. This is defined as follows:

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \tag{19}$$

This can be interpreted with the help of a Venn Diagram as shown in Figure 6. Here, we see that the conditional probability, denoted by $P(E_1|E_2)$ is defined by the ratio of the probability of the intersection of the events E_1 and E_2 (dark shaded region) and the probability of event E_2 (entire shaded region, light and dark).

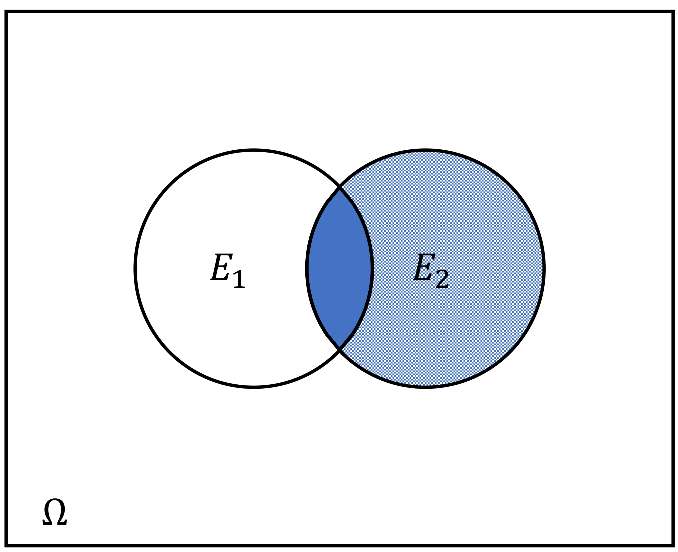

Figure 6: Venn Diagram illustrating conditional probability

By rearranging the terms in Eq. (19) we can express the probability of the intersection of two events, E_1 and E_2 as

$$P(E_1 \cap E_2) = P(E_1|E_2)P(E_2) \tag{20}$$

Events E_1 and E_2 are said to be *independent* if

$$P(E_1|E_2) = P(E_1) \tag{21}$$

That is, the probability of observing event E_1 is not influenced by the observation of event E_2. This further implies

$$P(E_1 \cap E_2) = P(E_1)P(E_2) \tag{22}$$

Note that *independent* events are not the same as *mutually exclusive* events, in which case $P(E_1|E_2) = 0$.

4 Law of Total Probability & Bayes' Rule

Using conditional probabilities, we will next introduce two very important results from probability theory that are used extensively in uncertainty quantification: The Law of Total Probability and Bayes' Rule. We've already heard a great deal about Bayes' Rule. We'll now see its mathematical form. But, first we need to introduce the Law of Total Probability.

Law of Total Probability: Consider that the sample space Ω is divided into n mutually exclusive events, E_1, E_2, $\ldots$, E_n. Let A also be an event in Ω. The probability of event A can be written as

$$\begin{aligned} P(A) &= P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_n) \\ &= P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_n)P(E_n) \\ &= \sum_{i=1}^{n} P(A|E_i)P(E_i) \end{aligned} \tag{23}$$

This is illustrated in Figure 7, which shows a sample space divided into eight mutually exclusive events $E_1, \ldots, E_8$. Also shown is another event A. Here, we can see that the probability of event A can be partitioned into the eight components that intersect the events $E_1, \ldots, E_8$.

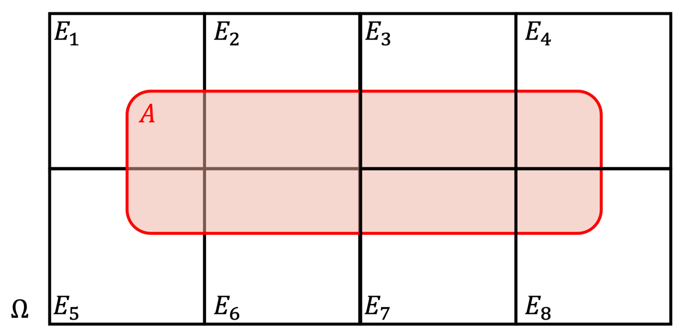

Figure 7: Venn Diagram demonstrating the Law of Total Probability

From Eq. (20) and the commutative law of intersections, we see that

$$P(A|E_i)P(E_i) = P(E_i|A)P(A) \tag{24}$$

which can be rewritten as

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{P(A)} \tag{25}$$

This is known as **Bayes' Rule**. By plugging this in the *Law of Total Probability*, Bayes' Rule can be expressed as

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum_{i=1}^{n} P(A|E_i)P(E_i)} \tag{26}$$

In light of our previous discussion of the Bayesian interpretation of probability, the components of Bayes' Rule have very useful interpretations:

- **Prior Probability**: $P(E_i)$ represents our belief in the probability of event E_i before observing A.
- **Likelihood**: $P(A|E_i)$ is the probability of observing event A given that event E_i occurs.
- **Posterior Probability:** $P(E_i|A)$ is the updated probability of event E_i after the event A has been observed.

References

[1] S. Ferson, L. R. Ginzburg, Different methods are needed to propagate ignorance and variability, Reliability Engineering & System Safety 54 (2-3) (1996) 133–144.

[2] M. H. Faber, On the treatment of uncertainties and probabilities in engineering decision analysis (2005).

[3] A. Der Kiureghian, O. Ditlevsen, Aleatory or epistemic? does it matter?, Structural safety 31 (2) (2009) 105–112.

[4] P. Walley, Statistical reasoning with imprecise probabilities (1991).

[5] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, in: Classic works of the Dempster-Shafer theory of belief functions, Springer, 2008, pp. 57–72.

[6] G. Shafer, A mathematical theory of evidence, Vol. 42, Princeton university press, 1976.

[7] D. Dubois, H. Prade, Possibility theory: an approach to computerized processing of uncertainty, Springer Science & Business Media, 2012.

[8] L. A. Zadeh, Fuzzy sets, Information and control 8 (3) (1965) 338–353.

[9] E. Hüllermeier, W. Waegeman, Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods, Machine Learning 110 (3) (2021) 457–506.

[10] S. Ferson, C. A. Joslyn, J. C. Helton, W. L. Oberkampf, K. Sentz, Summary from the epistemic uncertainty workshop: consensus amid diversity, Reliability Engineering & System Safety 85 (1-3) (2004) 355–369.

[11] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, Mechanical systems and signal processing 37 (1-2) (2013) 4–29.

Nomenclature

Functions

$P(\cdot|\cdot)$ Conditional probability of the event E_1 given the event E_2 has occurred

$P(\cdot)$ Probability measure

Operators

$\cap$ Intersection

$\cup$ Union

Variables

O Outcome from an experiment

E Event

E^* The complement of the event E

Ω Sample space

$\mathcal{F}$ Event space

$(\Omega, \mathcal{F}, P)$ Probability space, also known as a probability triple

$\emptyset$ The empty set, also known as the null set