Chapter 2 : Axioms of Probability

Notations.
- \(\mathcal{P}(S)\) (powerset of \(S\)) is the set of all subsets of \(S\)
- The relative complement of \(A\) in \(S\), denoted \(S \setminus A = \{ x \in S \mid x \notin A \}\).
  When the reference set \(S\) is clearly stated, \(S \setminus A\) may be simply denoted \(A^c\) and be called the complement of \(A\).
- A set \(S\) is said to be countable if there is a one-to-one correspondence between \(S\) and a subset of \(\mathbb{N}\)
- If \(S\) is a countable set, then we denote by \((S_i)_{i \in S}\) a countable sequence of sets indexed by \(S\)

1 Probability Space

We start by introducing mathematical concept of a probability space.

1.1 Sample Space and Events

Probability theory is mainly concerned with random experiments.

Definition 1.1 (Random experiment - Sample space). A random experiment is a phenomenon whose outcome is not predictable with certainty, but the set of all possible outcomes is known. The set of all possible outcomes is known as the sample space of the experiment and is denoted by \(\Omega\).

Definition 1.2 (Event). An event \(E\) is a set consisting of possible outcomes of the experiment that satisfy a given property. Thus, \(E\) is a subset of \(\Omega\) (\(E \in \mathcal{P}(\Omega)\)). If the outcome of the experiment is contained in \(E\), then we say that \(E\) has been realized or that \(E\) has occurred.

Any event \(E\) is a subset of \(\Omega\). Assume that the set of all events are represented by a particular family of sets over \(\Omega\) denoted \(\mathcal{A}\), i.e. \(\mathcal{A} \subseteq \mathcal{P}(\Omega)\). Which “desirable” properties should \(\mathcal{A}\) satisfy?

Since an event is defined as a set, let’s review basic operations on events.

Operations on events.
- \(\Omega\) is referred to as the sure event.
- \(\emptyset\) is referred to as the impossible event.
• Union: For any two events $E$ and $F$, the event $E \cup F$ consists of all outcomes that are either in $E$ or in $F$, meaning that $E \cup F$ is realized if either $E$ or $F$ occurs.

• Countable union: If $(E_i)_{i \geq 1}$ is a countable sequence of events, the union of these events denoted $\bigcup_{i=1}^{\infty} E_i$ is defined to be that event which consists of all outcomes that are in $E_i$ for at least one value of $i \in \mathbb{N}$, $i \geq 1$.

• Intersection: For any two events $E$ and $F$, the event $E \cap F$ consists of all outcomes that are both in $E$ and in $F$, meaning that $E \cap F$ is realized if both $E$ and $F$ occur.

• Countable intersection: If $(E_i)_{i \geq 1}$ is a countable sequence of events, the intersection of these events denoted $\bigcap_{i=1}^{\infty} E_i$ is defined to be that event which consists of all outcomes that are in all of the events $E_i$, $i \in \mathbb{N}$, $i \geq 1$.

• $E$ and $F$ are said to be mutually exclusive if $E \cap F = \emptyset$: $E \cap F$ is the impossible event, meaning that $E$ and $F$ cannot both occur in the same time.

• For any event $E$, we define the event $E^c$, referred to as the complement of $E$, to consist of all outcomes in the sample space $\Omega$ that are not in $E$, meaning that $E^c$ is realized if $E$ does not occur. Note that $E \cap E^c = \emptyset$ and $E \cup E^c = \Omega$.

A graphical representation that is useful for illustrating relations among events is the Venn diagram. The sample space is represented by a large rectangle, events are represented by circles and events of interests are indicated by shading the appropriate regions of the diagram.

**Reminder : Laws on sets.** For any three events $E$, $F$ and $G$

• Commutativity: $E \cup F = F \cup E$ and $E \cap F = F \cap E$.

• Associativity: $E \cup (F \cup G) = (E \cup F) \cup G$ and $E \cap (F \cap G) = (E \cap F) \cap G$.

• Distributivity: $E \cup (F \cap G) = (E \cup F) \cap (E \cup G)$ and $E \cap (F \cup G) = (E \cap F) \cup (E \cap G)$.

• Transitivity: If $E \subseteq F$ and $F \subseteq G$ then $E \subseteq G$.

• De Morgan’s laws: for any countable sequence of events $(E_i)_{i \geq 1}$:

\[
\left( \bigcup_{i=1}^{\infty} E_i \right)^c = \bigcap_{i=1}^{\infty} E_i^c,
\]

\[
\left( \bigcap_{i=1}^{\infty} E_i \right)^c = \bigcup_{i=1}^{\infty} E_i^c.
\]
Here, the event space is modelled as a \(\sigma\)-algebra on \(\Omega\).

**Definition 1.3 (\(\sigma\)-algebra).** A \(\sigma\)-algebra \(\mathcal{A} \subseteq \mathcal{P}(\Omega)\) is a family of sets over \(\Omega\) satisfying the following properties:

1. \(\emptyset \in \mathcal{A}\)
2. \(\mathcal{A}\) is closed under complementation: If \(A \in \mathcal{A}\), then \(A^c \in \mathcal{A}\)
3. \(\mathcal{A}\) is closed under countable union: If \((A_i)_{i \geq 1}\) is a countable sequence of sets in \(\mathcal{A}\) (i.e. \(A_i \in \mathcal{A}\) for \(i \in \mathbb{N}, i \geq 1\)), then \(\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}\)

**Definition 1.4 (Measurable space).** Let us consider a random experiment with sample space \(\Omega\) endowed with \(\sigma\)-algebra \(\mathcal{A}\), \((\Omega, \mathcal{A})\) is called a measurable space and elements of \(\mathcal{A}\) are called events.

### 1.2 Axioms of probability

To avoid any philosophical debate on randomness, we introduce an axiom system proposed by Kolmogorov that quickly became the mostly undisputed basis for modern probability theory.

**Definition 1.5 (Probability. Kolmogorov, 1933).** Let \((\Omega, \mathcal{A})\) be a measurable space of events. A probability measure is a real-valued function mapping \(\mathbb{P} : \mathcal{A} \to \mathbb{R}\) satisfying:

1. for any event \(E \in \mathcal{A}\), \(\mathbb{P}(E) \geq 0\) [Nonnegativity of the probability measure]
2. \(\mathbb{P}(\Omega) = 1\)
3. for any countably infinite sequence of events \((E_i)_{i \geq 1}\) that are mutually exclusive (i.e. \(E_i \cap E_j = \emptyset\) if \(i \neq j\)),

\[
\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)
\]

Axiom 3’s property is called \(\sigma\)-additivity or countable additivity

Then \((\Omega, \mathcal{A}, \mathbb{P})\) is called a probability space.

### 1.3 Some simple Properties

Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space.

**Property 1.1 (Probability of the impossible event).**

\[
\mathbb{P}(\emptyset) = 0
\]  \(\quad (1)

**Proof.** Let \((E_i)_{i \geq 1}\) be a countable sequence of events defined as follows:
• \( E_1 = \Omega \)
• \( E_i = \emptyset \) for all \( i > 1 \)

Since \( (E_i)_{i \geq 1} \) is a family of mutually exclusive events (\( \Omega \cap \emptyset = \emptyset \) and \( \emptyset \cap \emptyset = \emptyset \)), we have

\[
1 = P(\Omega) = P\left( \bigcup_{i=1}^{\infty} E_i \right) = \sum_{i=1}^{\infty} P(E_i) = P(\Omega) + \sum_{i=2}^{\infty} P(\emptyset)
\]

Thus, \( \sum_{i=2}^{\infty} P(\emptyset) = 0 \). Axiom (1) states that \( P(\emptyset) \geq 0 \). If \( P(\emptyset) > 0 \), then \( \sum_{i=2}^{\infty} P(\emptyset) = \infty \). Therefore \( P(\emptyset) = 0 \).

**Property 1.2** (Probability of a finite union of mutually exclusive events). For any finite sequence of events \( E_1, \ldots, E_n \in \mathcal{A} \) that are mutually exclusive (i.e. \( E_i \cap E_j = \emptyset \) if \( i \neq j \)),

\[
P\left( \bigcup_{i=1}^{n} E_i \right) = \sum_{i=1}^{n} P(E_i)
\]  

**Proof.** Let us consider \( (E_i)_{i \geq 1} \) a countably infinite sequence of mutually exclusive events such that \( E_i = \emptyset \) for all \( i > n \). We have that

\[
P\left( \bigcup_{i=1}^{n} E_i \right) = P\left( \bigcup_{i=1}^{\infty} E_i \right) = \sum_{i=1}^{\infty} P(E_i) = \sum_{i=1}^{n} P(E_i) + \sum_{i=n+1}^{\infty} P(\emptyset) = \sum_{i=1}^{n} P(E_i)
\]

**Property 1.3** (Probability of included events). For any two events \( E, F \in \mathcal{A} \),

\[
E \subseteq F \Rightarrow P(E) \leq P(F)
\]  

**Proof.** Using a Venn diagram representation, one might notice that if \( E \subseteq F \), then \( F = E \cup (F \cap E^c) \). Since \( E \) and \( F \cap E^c \) are mutually exclusive, we obtain, from Axiom 3 : \( P(F) = P(E) + P(F \cap E^c) \), thereby completing the proof since \( P(F \cap E^c) \geq 0 \).

**Corollary 1.1.** For any event \( E \in \mathcal{A} \),

\[
P(E) \leq 1
\]  

**Proof.** By definition, any event \( E \) is a subset of the sample space \( \Omega : E \subseteq \Omega \). Therefore, \( P(E) \leq P(\Omega) = 1 \).

**Property 1.4** (Law of total probability). Let \( F \in \mathcal{A} \) be an event and \( (E_i)_{i \geq 1} \) be a countable partition of the sample space \( \Omega \) (i.e. \( \bigcup_{i=1}^{\infty} E_i = \Omega \) and \( E_i \cap E_j = \emptyset \) for \( i \neq j \)),

\[
P(F) = \sum_{i=1}^{\infty} P(F \cap E_i)
\]
Proof.

\[
F = F \cap \Omega = F \cap \left( \bigcup_{i=1}^{\infty} E_i \right) = \bigcup_{i=1}^{\infty} (F \cap E_i)
\]

\((F \cap E_i)_{i \geq 1}\) is a countable sequence of mutually exclusive events. Thus, by using axiom (3)

\[
P(F) = P\left( \bigcup_{i=1}^{\infty} (F \cap E_i) \right) = \sum_{i=1}^{\infty} P(F \cap E_i)
\]

\[\square\]

**Property 1.5** (Probability of the complement). For any event \(E \in \mathcal{A}\),

\[
P(E^c) = 1 - P(E)
\]

Proof. Let \(E \in \mathcal{A}\) be an event. \(E\) and \(E^c\) form a partition of the sample space \(\Omega : \Omega = E \cap E^c\) and \(E \cap E^c = \emptyset\). According to the law of total probability,

\[
1 = P(\Omega) = P(\Omega \cap E) + P(\Omega \cap E^c) = P(E) + P(E^c)
\]

\[\square\]

**Property 1.6** (Probability of the union of 2 arbitrary events). For any two events \(E, F \in \mathcal{A}\),

\[
P(E \cup F) = P(E) + P(F) - P(E \cap F)
\]

(6)

Proof. Using a Venn diagram representation to get some intuition, we can write \(E \cup F\) as the union of mutually exclusive events \(F\) and \(E \setminus F^c\). Therefore, \(P(E \cup F) = P(F) + P(E \cap F^c)\). According to the law of total probability, \(P(E) = P(E \cap F) + P(E \cap F^c)\). Hence, the result holds. \[\square\]

**Property 1.7** (Inclusion-Exclusion Identity/ Poincaré’s formula). For any finite sequence of events \(E_1, \ldots, E_n \in \mathcal{A}\),

\[
P\left( \bigcup_{i=1}^{n} E_i \right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(E_{i_1} \cap \cdots \cap E_{i_k})
\]

(7)

where \(\sum_{1 \leq i_1 < \cdots < i_k \leq n}\) means the sum for all subsets of \(\{1, \ldots, n\}\) of size \(k\)

Proof. By induction. \[\square\]

Example: For \(n = 3\), the inclusion-exclusion identity turns out to be :

\[
P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3)
\]

\[
- P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3)
\]

\[
+ P(E_1 \cap E_2 \cap E_3)
\]
2 Uniform Probability Measure on Finite Sample Spaces

Let $\Omega$ be a finite sample space: $\Omega = \{\omega_1, \ldots, \omega_n\}$ with $|\Omega| = n \in \mathbb{N}, n \geq 1$. In that case, any subset of $\Omega$ will be an event, meaning that we will always consider the $\sigma$-algebra $\mathcal{A} = \mathcal{P}(\Omega)$ for a finite sample space $\Omega$. A probability measure $\mathbb{P}$ on measurable space $(\Omega, \mathcal{P}(\Omega))$ is fully characterized by the values $\mathbb{P}$ takes on outcomes $\omega_i$. Indeed an event $E$ can be written as:

$$E = \bigcup_{i \in J} \{w_i\}$$

where $J \subseteq \{1, \ldots, n\}$ is the set of indices of all the outcomes $w_i$ that compose event $E$. Then, for any event $E$,

$$\mathbb{P}(E) = \mathbb{P}\left(\bigcup_{i \in J} \{\omega_i\}\right) = \sum_{i \in J} \mathbb{P}(\{\omega_i\})$$

**Definition 2.1** (Uniform Probability Measure). Let $\Omega$ be a finite sample space: $\Omega = \{\omega_1, \ldots, \omega_n\}$ with $|\Omega| = n \in \mathbb{N}, n \geq 1$ and $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ be a probability space. Probability measure $\mathbb{P}$ is said to be **uniform** if all outcomes $\omega_i$ in the sample space are equally likely to occur, i.e. $\mathbb{P}(\{\omega_i\}) = \alpha$, for $i = 1, \ldots, n$, with $\alpha \geq 0$.

**Property 2.1.** Let $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ be a probability space with uniform probability measure $\mathbb{P}$ on finite sample space $\Omega = \{\omega_1, \ldots, \omega_n\}$. Then

$$\mathbb{P}(\{\omega_i\}) = \frac{1}{n}, \text{ for } i = 1, \ldots, n$$

**Proof.** Elementary events $\{\omega_1\}, \ldots, \{\omega_n\}$ form a partition of $\Omega$. According to the law of total probability,

$$1 = \mathbb{P}(\Omega) = \sum_{i=1}^{n} \mathbb{P}(\{\omega_i\}) = n\alpha$$

thereby completing the proof \(\Box\)

**Property 2.2.** Let $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ be a probability space with uniform probability measure $\mathbb{P}$ on finite sample space $\Omega = \{\omega_1, \ldots, \omega_n\}$. Then, for any event $E$

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|}$$

**Proof.** We already saw that $\mathbb{P}(E) = \sum_{i \in J} \mathbb{P}(\{\omega_i\})$ where $J \subseteq \{1, \ldots, n\}$ is the set of indices of all the outcomes $w_i$ that compose event $E$. Since all outcomes are equally likely to occur, we obtain

$$\mathbb{P}(E) = \sum_{i \in J} \frac{1}{n} = \frac{|E|}{|\Omega|}$$

\(\Box\)