MTL108
Probability function (measure)

Rahul Singh

Definition 1 (Algebra or field)**.** An algebra on a non-empty set Ω is a collection $\mathcal{C}$ of subsets of Ω satisfying:

1. $\emptyset \in \mathcal{C}$ and $\Omega \in \mathcal{C}$.
2. If $A \in \mathcal{C}$, then $A^c \in \mathcal{C}$ (closed under complementation).
3. If $A_1, A_2, \ldots, A_n \in \mathcal{C}$, then $A_1 \cup A_2 \cup \ldots \cup A_n \in \mathcal{C}$ (closed under finite unions).

Definition 2 (σ-algebra (field))**.** A σ-algebra on a non-empty set Ω is a collection $\mathcal{C}$ of subsets of Ω satisfying:

1. $\emptyset \in \mathcal{C}$ and $\Omega \in \mathcal{C}$.
2. If $A \in \mathcal{C}$, then $A^c \in \mathcal{C}$ (closed under complementation).
3. If $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{C}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$ (closed under countable unions).

Definition 3. A σ-algebra on a sample space is also known as a collection of events.

Example 1. Let $\Omega = \mathbb{N} = \{1, 2, 3, \ldots\}$. We define a collection of subsets $\mathcal{F}$ as

$$\mathcal{F} = \{A \subseteq \mathbb{N} : A \text{ is finite or } A^c \text{ is finite}\}$$

Sets whose complements are finite are often called **cofinite** sets.

Why $\mathcal{F}$ is a Field

A collection $\mathcal{F}$ is a field if it satisfies three properties:

- **Non-emptiness:** $\emptyset$ is finite, so $\emptyset \in \mathcal{F}$.
- **Closure under Complements:** If $A \in \mathcal{F}$, then by definition either A is finite (making A^c cofinite) or A^c is finite. In either case, $A^c \in \mathcal{F}$.
- **Closure under Finite Unions:** If $A, B \in \mathcal{F}$, their union $A \cup B$ is in $\mathcal{F}$.
  - If both A and B are finite, $A \cup B$ is finite.
  - If at least one (say A) is cofinite, then $(A \cup B)^c = A^c \cap B^c$. Since A^c is finite, the intersection $A^c \cap B^c$ must be finite, making $A \cup B$ cofinite.

Why $\mathcal{F}$ is NOT a σ-field

To be a σ-field, $\mathcal{F}$ must be closed under **countable** unions. Consider the sequence of singleton sets:

$$A_n = \{2n\} \quad \text{for } n = 1, 2, 3, \ldots$$

Each $A_n \in \mathcal{F}$ because every singleton set is finite. Now, consider the countable union:

$$A = \bigcup_{n=1}^{\infty} A_n = \{2, 4, 6, \ldots\}$$

The set A (the set of all even numbers) is **infinite**. Its complement, $A^c = \{1, 3, 5, \ldots\}$ (the set of all odd numbers), is also **infinite**.
Since neither A nor A^c is finite, $A \notin \mathcal{F}$. Thus, $\mathcal{F}$ is not closed under countable unions and is therefore not a σ-field.

Definition 4 (Probability function /measure)**.** Let Ω be a non-empty set and $\mathcal{C}$ be a σ-algebra on Ω. A probability space a triplet $(\Omega, \mathcal{C}, \mathbb{P})$, where function $\mathbb{P} : \mathcal{C} \to [0, 1]$ satisfies the following axioms:

1. $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\emptyset) = 0$
2. If $A_1, A_2, \ldots$ are disjoint events, that is, $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$\mathbb{P}(A_1 \cup A_2 \cup \ldots) = \sum_{k \geq 1} \mathbb{P}(A_k)$$

A probability space is a mathematical model for random phenomena, defined as a triple $(\Omega, \mathcal{C}, P)$, where Ω is the sample space (all possible outcomes), $\mathcal{C}$ is a σ-algebra of events (subsets of Ω), and $P : \mathcal{C} \to [0, 1]$ is a probability measure satisfying $P(\emptyset) = 0$, $P(\Omega) = 1$, and countable additivity for disjoint events.

Remark 1. The elements of $\mathcal{C}$ are events in a context. The function $\mathbb{P}$ takes an event $A \subset \omega$ as input and returns $\mathbb{P}(A)$, a real number between 0 and 1, as output.

Remark 2. Probability measure extends probability beyond equally likely notion.

Remark 3. In probability theory, the fair coin toss and six-sided die roll exemplify discrete uniform distributions on finite sample spaces. The coin assigns equal probability $1/2$ to heads or tails, modeling binary decisions, while the die distributes $1/6$ across outcomes $\{1, 2, 3, 4, 5, 6\}$, illustrating equiprobable events. Both underscore the importance of well-defined probability spaces for rigorous statistical inference.

Remark 4 (Axioms of probability)**.** The defining properties of a probability measure are usually referred to as the three axioms of probability, these are:

Axiom 1 (Non-negativity): $\mathbb{P}(A) \geq 0$

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

Axiom 3 (Countable Additivity): For a collection of pairwise disjoint events $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{C}$ (i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$),

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

> **Example 2.** Fair Coin Toss: $\Omega = \{H, T\}$ (heads or tails). $\mathcal{C} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$ (power set, a σ-algebra). $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\{H\}) = P(\{T\}) = \frac{1}{2}$, $\mathbb{P}(\{H, T\}) = 1$. This models equal chance for heads or tails.

> **Example 3.** Fair Six-Sided Die Roll: $\Omega = \{1, 2, 3, 4, 5, 6\}$. $\mathcal{C} = 2^{\Omega}$ (the power set: all $2^6 = 64$ subsets of Ω, forming a σ-algebra; that is,
>
> $$2^{\Omega} = \left\{ \begin{array}{c} \emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \\ \{1,2\}, \{1,3\}, \{1,4\}, \{1,5\}, \{1,6\}, \{2,3\}, \{2,4\}, \{2,5\}, \{2,6\}, \{3,4\}, \{3,5\}, \{3,6\}, \\ \{4,5\}, \{4,6\}, \{5,6\}, \\ \{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \{1,2,6\}, \{1,3,4\}, \{1,3,5\}, \{1,3,6\}, \{1,4,5\}, \{1,4,6\}, \\ \{1,5,6\}, \{2,3,4\}, \{2,3,5\}, \{2,3,6\}, \{2,4,5\}, \{2,4,6\}, \{2,5,6\}, \{3,4,5\}, \{3,4,6\}, \\ \{3,5,6\}, \{4,5,6\}, \\ \{1,2,3,4\}, \{1,2,3,5\}, \{1,2,3,6\}, \{1,2,4,5\}, \{1,2,4,6\}, \{1,2,5,6\}, \{1,3,4,5\}, \\ \{1,3,4,6\}, \{1,3,5,6\}, \{1,4,5,6\}, \{2,3,4,5\}, \{2,3,4,6\}, \{2,3,5,6\}, \\ \{2,4,5,6\}, \{3,4,5,6\}, \\ \{1,2,3,4,5\}, \{1,2,3,4,6\}, \{1,2,3,5,6\}, \{1,2,4,5,6\}, \{1,3,4,5,6\}, \{2,3,4,5,6\}, \\ \{1,2,3,4,5,6\} \end{array} \right\}$$
>
> For $\mathbb{P}(\{k\}) = \frac{1}{6}$ for each $k \in \Omega$; for event $A \subseteq \Omega$, $\mathbb{P}(A) = \frac{|A|}{6}$. E.g., probability of even number: $\mathbb{P}(\{2, 4, 6\}) = \frac{3}{6} = \frac{1}{2}$.

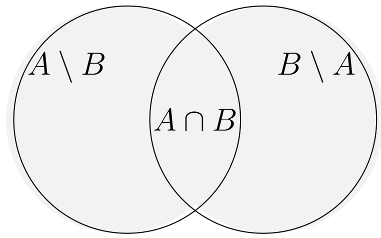
Properties

Theorem 1. *For any two events A and B and sample space Ω. We have*

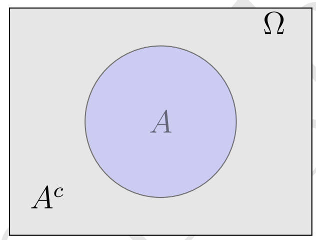
1. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
2. $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$, $A^c = \Omega \setminus A$.
3. *If* $A \subseteq B$ *then* $\mathbb{P}(A) \leq \mathbb{P}(B)$.
4. $0 \leq \mathbb{P}(A) \leq 1$.
5. $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$

Proof. Let $(\Omega, \mathcal{C}, \mathbb{P})$ be a probability space, where Ω is the sample space, $\mathcal{C}$ is a σ-algebra, and $\mathbb{P} : \mathcal{C} \to [0, 1]$ is a probability measure.

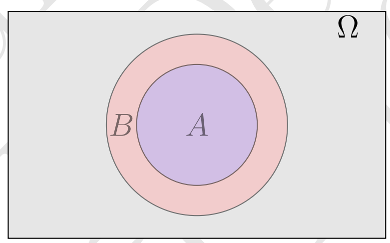
1. We can express $A \cup B = A \cup (B \setminus A)$, where A and $B \setminus A = B \cap A^c$ are disjoint. By the countable additivity axiom, $\mathbb{P}(A \cup (B \setminus A)) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$. Now, $B = (B \cap A) \cup (B \cap A^c)$, and since $B \cap A$ and $B \cap A^c$ are disjoint, $\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c)$. Thus, $\mathbb{P}(B \setminus A) = \mathbb{P}(B \cap A^c) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$. Therefore, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.



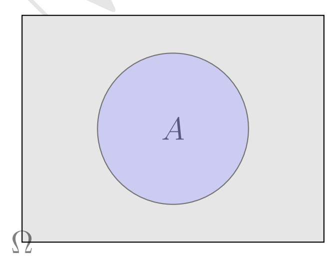
2. Since $A \cup A^c = \Omega$ and $A \cap A^c = \emptyset$, the events A and A^c are disjoint. By the countable additivity axiom, $\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$. Since $\mathbb{P}(\Omega) = 1$, we have $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$.



3. If $A \subseteq B$, then $B = A \cup (B \setminus A)$, where A and $B \setminus A$ are disjoint. By countable additivity, $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$. Since $\mathbb{P}(B \setminus A) \geq 0$, it follows that $\mathbb{P}(B) \geq \mathbb{P}(A)$.



4. Since $\mathbb{P}$ is a probability measure, by definition, for any event $A \in \mathcal{C}$, $\mathbb{P}(A) \geq 0$. Also, $\mathbb{P}(\Omega) = 1$, and since $A \subseteq \Omega$, $\mathbb{P}(A) \leq \mathbb{P}(\Omega) = 1$ by monotonicity (see part 3). Thus, $0 \leq \mathbb{P}(A) \leq 1$.



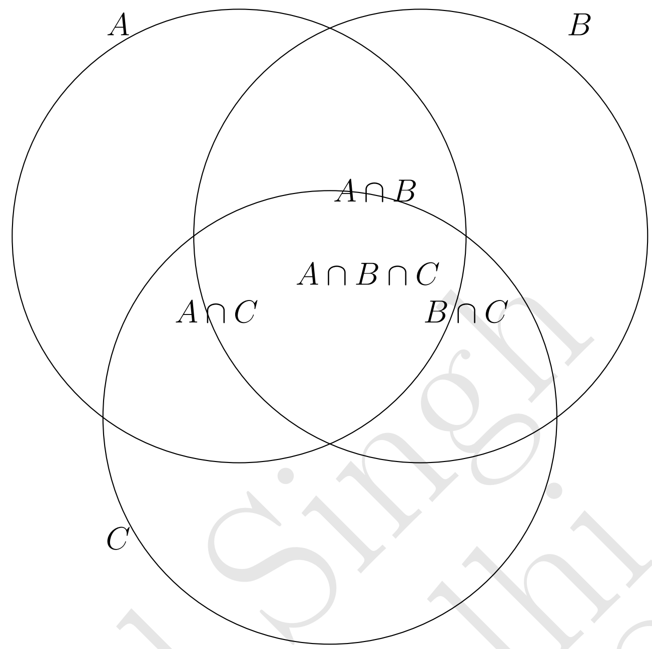
5. Follows from part 2 and non-negativity.

□

Theorem 2. *For any three events A, B and C,*

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C). \quad (1)$$

Proof. Consider the union $A \cup B \cup C$. The naive sum $\mathbb{P}(A)+\mathbb{P}(B)+\mathbb{P}(C)$ counts the probabilities of each event, but overcounts the intersections. A Venn diagram (see below) shows three circles representing A, B, and C, with overlapping regions.

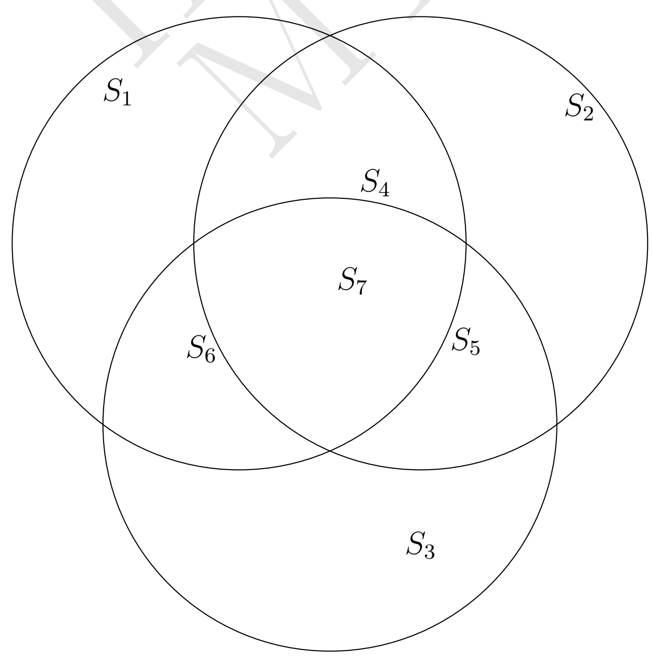


The pairwise intersections $A \cap B$, $A \cap C$, and $B \cap C$ are subtracted to correct double-counting. However, the triple intersection $A \cap B \cap C$ (central region) is subtracted thrice (once per pair) and must be added back once. Thus:

$$\mathbb{P}(A \cup B \cup C) = [\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)] - [\mathbb{P}(A \cap B) + \mathbb{P}(A \cap C) + \mathbb{P}(B \cap C)] + \mathbb{P}(A \cap B \cap C).$$

□

Aliter: Decompose $A \cup B \cup C$ into disjoint regions: $S_1 = A \setminus (B \cup C)$, $S_2 = B \setminus (A \cup C)$, $S_3 = C \setminus (A \cup B)$, $S_4 = (A \cap B) \setminus C$, $S_5 = (A \cap C) \setminus B$, $S_6 = (B \cap C) \setminus A$, and $S_7 = A \cap B \cap C$. The representation, via, Venn diagram, is as follows:



Note that sets $S_1, S_2, S_3, S_4, S_5, S_6, S_7$ are disjoint, so we have

$$\begin{aligned}
\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(S_1 \cup S_2 \cup S_3 \cup S_4 \cup S_5 \cup S_6 \cup S_7) = \mathbb{P}(S_1) + \mathbb{P}(S_2) + \mathbb{P}(S_3) + \ldots + \mathbb{P}(S_7) \\
\mathbb{P}(A) &= \mathbb{P}(S_1 \cup S_4 \cup S_6 \cup S_7) = \mathbb{P}(S_1) + \mathbb{P}(S_4) + \mathbb{P}(S_6) + \mathbb{P}(S_7) \\
\mathbb{P}(B) &= \mathbb{P}(S_2 \cup S_4 \cup S_5 \cup S_7) = \mathbb{P}(S_2) + \mathbb{P}(S_4) + \mathbb{P}(S_5) + \mathbb{P}(S_7) \\
\mathbb{P}(C) &= \mathbb{P}(S_3 \cup S_5 \cup S_6 \cup S_7) = \mathbb{P}(S_3) + \mathbb{P}(S_5) + \mathbb{P}(S_6) + \mathbb{P}(S_7) \\
\mathbb{P}(A \cap B) &= \mathbb{P}(S_4 \cup S_7) = \mathbb{P}(S_4) + \mathbb{P}(S_7) \\
\mathbb{P}(A \cap C) &= \mathbb{P}(S_6 \cup S_7) = \mathbb{P}(S_6) + \mathbb{P}(S_7) \\
\mathbb{P}(B \cap C) &= \mathbb{P}(S_5 \cup S_7) = \mathbb{P}(S_5) + \mathbb{P}(S_7) \\
\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(S_7).
\end{aligned}$$

So, the RHS of (1)

$$\begin{aligned}
&\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C) \\
&= \mathbb{P}(S_1) + \mathbb{P}(S_4) + \mathbb{P}(S_6) + \mathbb{P}(S_7) + \mathbb{P}(S_2) + \mathbb{P}(S_4) + \mathbb{P}(S_5) + \mathbb{P}(S_7) + \mathbb{P}(S_3) + \mathbb{P}(S_5) + \mathbb{P}(S_6) + \mathbb{P}(S_7) \\
&\quad - \Big(\mathbb{P}(S_4) + \mathbb{P}(S_7)\Big) - \Big(\mathbb{P}(S_6) + \mathbb{P}(S_7)\Big) - \Big(\mathbb{P}(S_5) + \mathbb{P}(S_7)\Big) + \mathbb{P}(S_7) \\
&= \mathbb{P}(S_1) + \mathbb{P}(S_2) + \mathbb{P}(S_3) + \mathbb{P}(S_4) + \mathbb{P}(S_5) + \mathbb{P}(S_6) + \mathbb{P}(S_7) \\
&= \mathbb{P}(A \cup B \cup C).
\end{aligned}$$

Theorem 3 (Total probability law)**.** *Let $\{A_1, A_2, \ldots, A_n\} \subset \mathcal{C}$ be a partition of Ω, that is, $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$ and $A_i \cap A_j = \emptyset$. Then, for any event B,*

$$\mathbb{P}(B) = \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \ldots + \mathbb{P}(B \cap A_n).$$

Proof. Since $\{A_1, A_2, \ldots, A_n\}$ is a partition, the events $B \cap A_i$ are disjoint (because $A_i \cap A_j = \emptyset$ for $i \neq j$ implies $(B \cap A_i) \cap (B \cap A_j) = B \cap (A_i \cap A_j) = B \cap \emptyset = \emptyset$). Their union covers B:

$$\bigcup_{i=1}^{n} (B \cap A_i) = B \cap \left(\bigcup_{i=1}^{n} A_i\right) = B \cap \Omega = B,$$

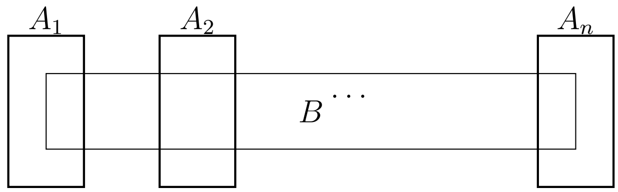
as $\bigcup_{i=1}^{n} A_i = \Omega$. By the countable additivity axiom, for disjoint events:

$$P\left(\bigcup_{i=1}^{n} (B \cap A_i)\right) = \sum_{i=1}^{n} P(B \cap A_i).$$

Since the union is B, we have:

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i).$$

A Venn diagram (adapted for partitions) shows B as the region across disjoint A_i (see below). Each $B \cap A_i$ is the part of B within A_i, and their sum equals $P(B)$.



This visualizes B as the union of $B \cap A_i$, supporting the additivity. $\square$

References

[1] Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability.* Chapman and Hall/CRC.

Disclaimer