Table of Contents

Notation

Throughout this summary, we are going to use the following notation:

- $\cup \longleftrightarrow$ OR
- $\cap \longleftrightarrow$ AND
- $A^c \longleftrightarrow$ NOT A
- $\omega \in A \longleftrightarrow \omega$ "belongs in" A
- $\omega \notin A \longleftrightarrow \omega$ "does not belong in" A

Sample Space, Events and Set Theory

Sample Space

The **sample space** Ω is the set of all the possible outcomes that can be produced by the random experiment. For example,

- if the random experiment is a single coin flip, then $\Omega = \{H, T\}$.
- if the random experiment is single six-sided die roll, then $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- if the random experiment is a consecutive toss of a coin, then $\Omega = \{HH, HT, TH, TT\}$.

From now on, we are going to consider that $\Omega = \{\omega : \omega \text{ simple event}\}$ is finite.

Events

An **event** A is any subset of Ω. This means that it contains a small collection of simple outcomes. For example, if the random experiment is a single six-sided die roll, then $\Omega = \{1, 2, 3, 4, 5, 6\}$ and some events could be:

- $A = \{1, 4\}$
- $A = \{ \text{ only the odd numbers } \} = \{1, 3, 5\}$
- $A = \Omega$
- $A = \emptyset$ (the empty set, it contains nothing)

Set Operations

There are three operations that can be applied in sets. Using these three operations and combining them, we can describe all the events of Ω.

Complement

If A is an event, then A^c $(\text{not } A)$ is the complement of A, and it is defined as

$$A^c = \{\text{"Everything that does not belong in A"}\} = \{\omega \in \Omega : \omega \notin A\}$$

For example, if the random experiment is a single six-sided die roll, then $\Omega = \{1, 2, 3, 4, 5, 6\}$ and we take the event $A = \{1, 2\}$, then $A^c = \{3, 4, 5, 6\}$.

Union

If A, B are two events, then $A \cup B$ (A or B) is the union of A and B, and it is defined as

$$A \cup B = \{\text{"Everything that is in A or B or both of them"}\} = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$$

For example, if the random experiment is a single six-sided die roll, then $\Omega = \{1, 2, 3, 4, 5, 6\}$. Let's take the events $A = \{1, 2\}$ and $B = \{3, 5\}$. Then $A \cup B = \{1, 2, 3, 5\}$.

Intersection

If A, B are two events, then $A \cap B$ (A or B) is the union of A and B, and it is defined as

$$A \cap B = \{\text{"Everything that is in A and B"}\} = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$$

For example, if the random experiment is a single six-sided die roll, then $\Omega = \{1, 2, 3, 4, 5, 6\}$. Let's take the events $A = \{1, 2, 3\}$ and $B = \{1, 3, 5, 6\}$. Then $A \cap B = \{1, 3\}$.

Set Difference

If A, B are two events, then $A \backslash B$ (A without B) is the difference of A minus B, and it is defined as

$$A \backslash B = \{\text{"Everything that is in A and not in B"}\} = \{\omega \in \Omega : \omega \in A \text{ and } \omega \notin B\}$$

Note: Set difference is not really a new operation, because it can be written as: $A \backslash B = A \cap B^c$

For example, if the random experiment is a single six-sided die roll, then $\Omega = \{1, 2, 3, 4, 5, 6\}$. Let's take the events $A = \{1, 3, 5, 6\}$ and $B = \{1, 2, 3\}$. Then $A \backslash B = \{5, 6\}$.

Properties of Set Operations

Distributive Property of Unions and Intersections

If we have three events A, B, C, then

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

and

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Note: Notice that this imitates the distributive propery of multiplication in the real numbers.

Laws for the Complement: De Morgan's Laws

If we have two events A, B, then

$$(A \cup B)^c = A^c \cap B^c$$

and

$$(A \cap B)^c = A^c \cup B^c$$

Note: Essentially, these rules suggest that in order to pass the complement inside the parenthesis, we have to exchange unions and intersections, and then pass the complement to each set.

This can be inductively extended to more than two sets:
If we have events $A_1, \dots, A_n$, with $n \geq 3$ then

$$(A_1 \cup \cdots \cup A_n)^c = A_1^c \cap \cdots \cap A_n^c$$

and

$$(A_1 \cap \cdots \cap A_n)^c = A_1^c \cup \cdots \cup A_n^c$$

Calculating Probabilities

Probability of an Event

If we want to compute the probability of an event A, then we need to find all the simple events that are included in A and then the probability of A will be:

$$P(A) = \sum_{\omega \in A} P(\{\omega\})$$

For example, if the random experiment is a single six-sided die roll, then $\Omega = \{1, 2, 3, 4, 5, 6\}$. If $A = \{\text{ only the odd numbers }\}$ then in order to find the probability of A we write:

$$A = \{\text{ only the odd numbers }\} = \{1, 3, 5\}$$

and so the probability is:

$$P(A) = P(\{1\}) + P(\{3\}) + P(\{5\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Disjoint sets and the Addition Rule

One key property is the behaviour of a probability with respect to disjoint sets. Two events A, B are called Disjoint (or Mutually Exclusive) events when $A \cap B = \varnothing$, which means that they do not have any common elements. In that case, we have that

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

This is what we call the "Addition Rule" and it can be extended inductively for more than 2 sets:
If $A_1, \dots, A_n$, with $n \geq 3$ are disjoint events (meaning that each one of them has no common elements with any of the other events),

$$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$$

The Complement Rule

For any event A, we have that A, A^c are disjoint and so using the addition rule $P(A \cup A^c) = P(\Omega) = P(A) + P(A^c)$. However, we know that $P(\Omega) = 1$. Therefore:

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

or equivalently

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

This rule is very useful because, sometimes, calculating the probability $P(A^c)$ is much easier than calculating the probability $P(A)$. If we find $P(A^c)$, then using the above, $P(A) = 1 - P(A^c)$

Law of Total Probability

If we have two events A, B, then we can write the probability of A as follows:

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

This way, we managed to write the probability $P(A)$ as the sum of the probabilities of two disjoint events. It is simple to derive this expression:

$$P(A) = P(A \cap \Omega) = P\Big(A \cap (B \cup B^c)\Big) = P\Big((A \cap B) \cup (A \cap B^c)\Big) = P(A \cap B) + P(A \cap B^c)$$

This can be extended to more than 2 sets:
If $B_1, \ldots, B_n$, with $n \geq 3$ are disjoint events **and** $B_1 \cup \cdots \cup B_n = \Omega$, then

$$P(A) = P(A \cap B_1) + \cdots + P(A \cap B_n)$$

General Rule for the Probability of a Union

In the case where we have events A, B that are not disjoint, then if we want to calculate the probability $P(A \cup B)$, we use the following:

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

Notice that in the case that A, B are disjoint, meaning that $A \cap B = \varnothing$, we have that $P(A \cap B) = 0$ and so we end up with the formula of the addition rule.

You can find examples and further details about the Addition rule in the [Data Science Discovery Lessons (https://discovery.cs.illinois.edu/learn/Prediction-and-Probability/Multi-event-Probability-Addition-Rule/)](https://discovery.cs.illinois.edu/learn/Prediction-and-Probability/Multi-event-Probability-Addition-Rule/) or in the [corresponding guide (https://discovery.cs.illinois.edu/guides/Statistics-Formulas/addition-rule/)](https://discovery.cs.illinois.edu/guides/Statistics-Formulas/addition-rule/).

Independent Events and the Multiplication Rule

Two events A, B are called independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

This is the definition of independent sets and this is also what we call the multiplication rule. It can be extended inductively to more than 2 events:
If $A_1, \ldots, A_n$, with $n \geq 3$ are independent events, then

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot \ldots \cdot P(A_n)$$

You can find examples and further details about Independent Events and the Multiplication Rule in the [Data Science Discovery Lessons (https://discovery.cs.illinois.edu/learn/Prediction-and-Probability/Multi-event-Probability-Multiplication-Rule/)](https://discovery.cs.illinois.edu/learn/Prediction-and-Probability/Multi-event-Probability-Multiplication-Rule/) or in the [corresponding guide (https://discovery.cs.illinois.edu/guides/Statistics-Formulas/multiplication-rule/)](https://discovery.cs.illinois.edu/guides/Statistics-Formulas/multiplication-rule/).

Conditional Probability

If we have two events A, B, we define the conditional probability of A given B as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

This expresion can be written equivalently as:

$$P(A \cap B) = P(A|B) \cdot P(B)$$

Remark: Notice that when the events A, B are disjoint, then $P(A|B) = 0$.

The second equation is the Multiplication Rule for non-Independent Events.

You can find examples and further details about the Conditional Probability in the [Data Science Discovery Lessons (https://discovery.cs.illinois.edu/learn/Prediction-and-Probability/Conditional-Probability/)](https://discovery.cs.illinois.edu/learn/Prediction-and-Probability/Conditional-Probability/).

Bayes rule

Suppose we have two events A, B. Then the Bayes Rule helps us calculate conditional probabilities by exchanging the roles of A and B:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Remark: The equation comes directly from the definition of the Conditional Probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$ and then we can write the numerator as follows $P(A \cap B) = P(B|A) \cdot P(A)$.

If we also apply the Law of Total Probability in the denominator, we can write:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$= \frac{P(B|A) \cdot P(A)}{P(B \cap A) + P(B \cap A^c)}$$

$$= \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c)}$$

All the expressions above are equivalent. So depending on the problem, we use the appropriate one.

You can find examples for the Bayes Rule in the Data Science Discovery Lessons