# Probability Theory

Daniela Rodriguez

# Schedule and Grading Policy

## Homework Schedule and Deadlines

Please note: **All dates listed are tentative and subject to change.** Any modifications will be announced promptly.
The dates listed below indicate when each homework assignment will become **available** on the course platform. The deadline for each assignment is the day immediately preceding the start date of the subsequent assignment. Please ensure you submit your work according to the due dates provided below:

| Assignment Topic | Available Date | Due Date |
|---|---|---|
| Homework 1: Probability Spaces | October 20 | October 29 |
| Homework 2: Conditional and Independence | October 30 | November 10 |
| Homework 3: Random Variables | November 11 | November 17 |
| Homework 4: Random Vectors | November 18 | December 3 |
| Homework 5: Expected Value | December 4 | December 17 |
| Homework 6: Convergence | December 18 | January 10 |

## Grading Policy

The final course grade (FC) is determined by the following policy, where $F$ is the Final Exam grade, $M$ is the Midterm grade, and $HW$ is the Homework grade (calculated as the average of your **best four HW scores**). The passing grade is 55.

Let $P = \max(M, F) \times 0.2 + F \times 0.8$ be the weighted passing threshold.

$$\text{FC} = \begin{cases} \max(HW, F) \times 0.15 + \max(M, F) \times 0.15 + F \times 0.7 & \text{if } P \geq 55 \\ \max(M, F) \times 0.2 + F \times 0.8 & \text{if } P < 55 \end{cases}$$

## Exam Dates

**Midterm Exam:** Date and time **To Be Determined (TBD)**.

**Exam A:**

- Date: January 19, 2026
- Time: 9:00

**Exam B:**

- Date: February 9, 2026
- Time: 9:00

## Bibliography


- **Ross, Sheldon M.** *A first course in probability.*
- **Feller, William.** *An introduction to probability theory and its applications.*
- **Hoel, Paul G.** *Introduction to probability theory.*
- **Grimmet, Geoffrey R.** *Probability and random processes.*

# Probability spaces

## Introduction

Consider the following experiment: a box has 4 red balls, 6 blue balls and 10 green balls. We pick a ball at random. What is the probability of that ball being red? We are taught in school that this should be the number of red balls over the total number of balls, so $\frac{4}{20} = 0.2 = 20\%$ and this is indeed true under certain assumptions. To understand the assumptions we are implicitly making when doing this computation, we ask the following questions:

What is a probability as a mathematical object?

Would the answer be the same if some balls are of difference sizes?

Note that when we ask about a probability, we need to determine the event whose probability we are interested in – while the probability of a specific event (e.g. 'the ball is red') is a number in $[0, 1]$, the probability on its own, is a map that attaches to each event a number. There are three possible outcomes for this experiment: red, blue and green.

The set of all elementary events is called a **Sample space** and is denoted by $\Omega$. Elements of $\Omega$, that is elementary events, are denoted by $\omega \in \Omega$.

As a mathematical object, $\Omega$ is any non-empty set – in this case, $\Omega = \{\text{red, blue, green}\}$

**Example 1.**

- *We throw a coin. There are two results: heads or tails. Thus* $\Omega = \{H,\ T\}$, $|\Omega| = 2$.
- *We throw a die. There are six possible results. Thus* $\Omega = \{1, 2, 3, 4, 5, 6\}$, $|\Omega| = 6$.

Often we are not interested in a concrete result of an experiment but we just want to know if this result belongs to a subset of $\Omega$. Such subsets are called **events** and we denote them with capital letters: $A, B, C, D$ etc.

**Example 2.**

- *We throw two dice. Let $A$ be an event that the sum of spots is equal to 5. Then* $\Omega = \{(i, j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6\}$, $A = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$.
- *We throw a coin until we get heads. Let $A$ be an event where at most 3 trials were done.* $\Omega = \{(H), (T,H), (T,T,H), (T,T,T,H), \dots\}$. $A = \{(H), (T,H), (T,T,H)\}$.

Since we'll be working with sets, we'll need to perform operations on them. So, let's review the notation.

**Remark 1.**

1. *$\Omega$ - the sample space*
2. *$\emptyset$ - the impossible event*
3. *$A \cap B$ - events $A$ and $B$ both occurred*
4. *$A \cap B = \emptyset$ - events $A$ and $B$ are mutually exclusive*
5. *$A \cup B$ - either $A$ or $B$ occurred*
6. *$A^c = \Omega \setminus A$ - $A$ did not happen*
7. *$A \setminus B = A \cap B^c$ - $A$ happened and $B$ did not happen*
8. *$A \subseteq B$ - event $A \neq \emptyset$ leads to event $B$*

For example, we can ask for the probability that 'the ball is either blue or green' (which would have been equivalent to 'ball is not red'). In words, an event is a statement that you can tell whether it is true or not, after seeing the outcome of the experiment. In this case, all possible events are

'The ball is none of the three colours or any other colour' – mathematically, this will be denoted by the empty set $\emptyset$, since it contains none of the possible outcomes.

'The ball is red' – denoted by $\{red\}$

'The ball is blue' – denoted by $\{blue\}$

'The ball is green' – denoted by $\{green\}$

'The ball is either red or blue' – denoted by $\{red, blue\}$

'The ball is either red or green' – denoted by $\{red, green\}$

'The ball is either blue or green' – denoted by $\{blue, green\}$

'The ball is any of red, blue or green' – denoted by $\{red, blue, green\} = \Omega$.

From this exhaustive list, it is clear that all events are subset of $\Omega$ and in fact, in this case at least, all subsets of $\Omega$ are events.

Assume we have a fixed sample space $\Omega$. We want to distinguish a family of events, that we want to consider. We call the collection of all events the event space, usually denoted by $\mathcal{F}$.

A first choice is to take: $2^\Omega =$ all subsets of $\Omega$. This is a good choice when $\Omega$ is at most countable set. When $|\Omega| > \aleph_0$, $2^\Omega$ is too big and there are problems with defining probability on $2^\Omega$. The set of all possible subsets of a given set is often denoted by $\mathcal{P}(A)$ or $2^\Omega$ and it is called the power set. We will discuss these problems later. That is why we need to choose a smaller family. On the other hand, $\mathcal{F}$ should be closed with respect to taking unions, intersections and complements.

**Definition 1.** *A family $\mathcal{F}$ of subsets of $\Omega$ is called a $\sigma$-**algebra** if:*

1. $\emptyset \in \mathcal{F}$
2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
3. $A_1, A_2, \dots, A_n, \dots \in \mathcal{F} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

*A pair $(\Omega, \mathcal{F})$ is called a **measurable space**.*

**Remark 2.** *For any* $\Omega$ *the pair* $(\Omega, 2^\Omega)$ *is a measurable space. Also for any nonempty subset* $A \subset \Omega$ *the smallest* $\sigma$*-algebra that contains* $A$ *is* $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$*.*

**Fact 1.** *Let* $(\Omega, \mathcal{F})$ *be a measurable space. Then*

1. $A, B \in \mathcal{F} \implies A \setminus B \in \mathcal{F}$
2. $A_1, A_2, \ldots, A_n, \cdots \in \mathcal{F} \implies \bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$

Proof.

1. $A \setminus B = A \cap B^c = (A^c \cup B)^c$
2. $\bigcap_{k=1}^{\infty} A_k = \left(\bigcup_{k=1}^{\infty} A_k^c\right)^c$, but $A_k^c \in \mathcal{F}$.

**Definition 2.** ***Borel*** $\sigma$***-algebra of*** $\mathbb{R}^d$, $\mathcal{B}(\mathbb{R}^d)$, *is the smallest* $\sigma$*-algebra that contains all open subsets of* $\mathbb{R}^d$*. Elements of* $\mathcal{B}(\mathbb{R}^d)$ *are called* ***Borel subsets****.*

**Example 3.** *Let $d = 1$, $\mathcal{B}(\mathbb{R})$ is the smallest $\sigma$-algebra that contains all open sets in $\mathbb{R}$.*

- *Intervals $(a, b)$, where $a, b \in \mathbb{R}$ are in $\mathcal{B}(\mathbb{R})$,*
- $(a, b] \in \mathcal{B}(\mathbb{R})$ *as* $(a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right)$,
- $[a, b) \in \mathcal{B}(\mathbb{R})$ *as* $[a, b) = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right)$,
- $[a, b] \in \mathcal{B}(\mathbb{R})$ *as* $[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right)$,
- $\{a\} \in \mathcal{B}(\mathbb{R})$ *as* $\{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a + \frac{1}{n}\right)$,
- $(-\infty, a) \in \mathcal{B}(\mathbb{R})$ *(or* $(-\infty, a) = \bigcup_{n=1}^{\infty}(-n, a)$*)* $\implies [a, \infty) \in \mathcal{B}(\mathbb{R})$,
- $(a, \infty) \in \mathcal{B}(\mathbb{R})$ *(or* $(a, \infty) = \bigcup_{n=1}^{\infty}(a, n)$*)* $\implies (-\infty, a] \in \mathcal{B}(\mathbb{R})$,
- $\mathbb{Q} \in \mathcal{B}(\mathbb{R})$ *as a countable union of points,*
- $\mathbb{R} \setminus \mathbb{Q} \in \mathcal{B}(\mathbb{R})$ *as the complement of* $\mathbb{Q}$.

**Remark 3.** *Not every subset of $\mathbb{R}^d$ is a Borel subset. An example of non-Borel set is a Vitali set $V \subset [0, 1]$.*

## Probability

We next define the notion of probability. First, however, in order to get some intuition we consider the **frequency of an event**. Assume that we can repeat an experiment $n$ times. Each repetitions happens under the same conditions. We define a relative frequency of an event $A \in \mathcal{F}$ in the series of $n$ experiments by:

$$f_n(A) = \frac{\#\text{ experiments in which } A \text{ happened}}{n}$$

When $n$ is large we expect that $f_n(A)$ is close to the chance $A$ occurs in a single trial. We easily check that $f_n$ takes values in $[0, 1]$ and

1. $f_n(\Omega) = 1$
2. If $A_1, \ldots, A_j$ are pairwise disjoint, then
$$f_n\left(\bigcup_{k=1}^{j} A_k\right) = \sum_{k=1}^{j} f_n(A_k).$$

This is because $\#$ experiments in which $\bigcup_{k=1}^{j} A_k$ happened is $\sum_{k=1}^{j}(\#$ experiments in which $A_k$ happened).

These are fundamental properties that a probability map should have. Are these sufficient for infinite probability spaces? Let us consider the following:

**Example 4.** *Let $\Omega = \mathbb{N}^* = \{1, 2, ...\}$ be the positive natural numbers and $\mathcal{F} = \mathcal{P}(\Omega)$. Suppose that $P(\{n\}) = \frac{1}{2^n}$, for every $n \geq 1$. What would we expect the event $\{2n | n \geq 1\}$ ('the outcome is an even number') to be?*

*Intuitively, what we would do is to sum up the probabilities corresponding to the outcome being even, i.e.,*

$$P(\{2n|n \geq 1\}) = \sum_{n=1}^{\infty} P(\{2n\}) = \sum_{n=1}^{\infty} \frac{1}{2^{2n}} = \sum_{n=1}^{\infty} \frac{1}{4^n}$$

*(Note that the event $\{n\}$ corresponds to 'the outcome is $n$'). The computation then follows as:*

$$\sum_{n=1}^{\infty} \frac{1}{4^n} = \frac{1}{1 - \frac{1}{4}} - 1 = \frac{1}{\frac{3}{4}} - 1 = \frac{1}{3}$$

The computation above cannot be justified, unless we extend the property of finite additivity to also hold for countable unions of disjoint events. Indeed, this is what we do!

**Definition 3.** *Given a sample space* $\Omega$ *and an event space* $\mathcal{F}$*, a function* $P : \mathcal{F} \to \mathbb{R}$ *is called a* probability measure *if it satisfies*

- $P(B) \in [0,1]$ *for every* $B \in \mathcal{F}$*;*
- $P(\Omega) = 1$*;*
- *(Countable additivity) For every* $A_n \in \mathcal{F}$*,* $n > 1$ *disjoint events (i.e. for all* $m, n >$ *such that* $m \neq n$*,* $A_m \cap A_n = \emptyset$*),*

$$P\left[\bigcup_{n=1}^{\infty} A_n\right] = \sum_{n=1}^{\infty} P(A_n).$$

We now give the definition of an abstract probability space.

**Definition 4.** *A* probability space *is defined as the triplet* $(\Omega, \mathcal{F}, P)$*, where*

- $\Omega$ *(the* sample space*) is the set of all possible outcomes of the experiment (we always assume that it is not empty);*
- $\mathcal{F}$ *is an event space of subsets of* $\Omega$*.*
- $P$ *is a probability measure on* $\mathcal{F}$*.*

**Theorem 1.** *Assume* $(\Omega, \mathcal{F}, P)$ *is a probability space, and* $A, B, A_1, \ldots, A_n, \cdots \in \mathcal{F}$*, then*

*1)* $P(\emptyset) = 0$

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

*3) If* $A \subseteq B$*, then* $P(B \setminus A) = P(B) - P(A)$ *and* $P(B) \geq P(A)$

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

*5)* $P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k)$

**Proof.**

1. $\Omega \cup \emptyset = \Omega$, $\Omega \cap \emptyset = \emptyset \implies 1 = P(\Omega) = P(\Omega) + P(\emptyset) \implies P(\emptyset) = 0$.
2. $\Omega = A \cup A^c$ and $A \cap A^c = \emptyset \implies 1 = P(\Omega) = P(A) + P(A^c) \implies P(A^c) = 1 - P(A)$.
3. $B = (B \setminus A) \cup A$ and $(B \setminus A) \cap A = \emptyset \implies P(B) = P(B \setminus A) + P(A)$.
4. $A \cup B = (A \setminus (A \cap B)) \cup (A \cap B) \cup (B \setminus (A \cap B))$.

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

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

5. Let $B_1 = A_1$, $B_2 = A_2 \setminus B_1$, $B_3 = A_3 \setminus (B_1 \cup B_2)$, $\ldots$, $B_n = A_n \setminus \bigcup_{k=1}^{n-1} B_k$. The $B_k$'s are mutually exclusive. $\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} B_k$, and $B_n \subset A_n$, for all $n$.

   Thus $P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} P(B_k) \leq \sum_{k=1}^{\infty} P(A_k)$.

We can use countable additivity to compute the probability of a union of disjoint events. How can we compute the probability of any union of events? The following proposition gives us a way to do this.

**Theorem 2** (Inclusion-Exclusion Formula)**.** *Assume $A_1, A_2, \ldots, A_n \in \mathcal{F}$, then:*

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k})$$

$$= \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

The sum $\sum_{1 \leq i_1 < \cdots < i_k \leq n}$ is to be interpreted as the sum going through all $k$-tuples $(i_1, \ldots, i_k)$ of numbers $\{1, \ldots, n\}$ with no repetition (inequalities are strict).

This Formula uses concise notation and is not straightforward to interpret. To understand it better, let us consider some specific cases.

For $n = 2$:

$$\begin{aligned} P\left[\bigcup_{k=1}^{2} A_k\right] &= \sum_{k=1}^{2} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq 2} P(A_{i_1} \cap \cdots \cap A_{i_k}) \\ &= \sum_{1 \leq i \leq 2} P(A_i) - \sum_{1 \leq i_1 < i_2 \leq 2} P(A_{i_1} \cap A_{i_2}) \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2). \end{aligned}$$

For $n = 3$:

$$
\begin{aligned}
P\left[\bigcup_{k=1}^{3} A_k\right] &= \sum_{k=1}^{3}(-1)^{k-1} \sum_{1\le i_1<\cdots<i_k\le 3} P(A_{i_1}\cap\cdots\cap A_{i_k}) \\
&= \sum_{1\le i\le 3} P(A_i) - \sum_{1\le i_1<i_2\le 3} P(A_{i_1}\cap A_{i_2}) + \sum_{1\le i_1<i_2<i_3\le 3} P(A_{i_1}\cap A_{i_2}\cap A_{i_3}) \\
&= P(A_1)+P(A_2)+P(A_3)-P(A_1\cap A_2)-P(A_1\cap A_3) \\
&\quad -P(A_2\cap A_3)+P(A_1\cap A_2\cap A_3).
\end{aligned}
$$

**Proof.** (by induction): By Theorem 1 we know that Theorem 2 is true for $n = 2$. Assume it is true for $n \geq 2$. We need to show that it is true for $n+1$. $P(A_1\cup A_2\cup\cdots\cup A_n\cup A_{n+1}) = P((A_1\cup A_2\cup\cdots\cup A_n)\cup A_{n+1})$
$= P(A_1\cup A_2\cup\cdots\cup A_n) + P(A_{n+1}) - P((A_1\cup A_2\cup\cdots\cup A_n)\cap A_{n+1})$
$= P(A_1\cup A_2\cup\cdots\cup A_n) + P(A_{n+1}) - P((A_1\cap A_{n+1})\cup\cdots\cup(A_n\cap A_{n+1}))$.
Now apply the induction hypothesis to $P(A_1\cup\cdots\cup A_n)$ and $P((A_1\cap A_{n+1})\cup\cdots\cup(A_n\cap A_{n+1}))$.
The full expansion becomes:

$$
P(A_1\cup A_2\cup\cdots\cup A_n\cup A_{n+1}) = \sum_{k=1}^{n}(-1)^{k-1}\sum_{1\le i_1<\cdots<i_k\le n} P(A_{i_1}\cap\cdots\cap A_{i_k}) + P(A_{n+1})
$$

$$
-\sum_{k=1}^{n}(-1)^{k-1}\sum_{1\le i_1<\cdots<i_k\le n} P(A_{i_1}\cap\cdots\cap A_{i_k}\cap A_{n+1})
$$

Combining these three formulas we get the result.

**Example 5.** ***Equiprobability.***: *$\Omega$ - a finite set, $\mathcal{F} = 2^{\Omega}$, all elementary events have the same probability. Then for all $\omega \in \Omega$ and for all $A \subseteq \Omega$ ($A \in \mathcal{F}$)*

$$P(\{\omega\}) = \frac{1}{|\Omega|} \tag{1}$$

$$P(A) = \frac{|A|}{|\Omega|} \tag{2}$$

*Since $\forall \omega_1, \omega_2 \in \Omega$, $P(\{\omega_1\}) = P(\{\omega_2\})$, let $p \in [0,1]$ such that $p := P(\{\omega\}) \forall \omega \in \Omega$.*
*Since $P$ is a probability measure*

$$1 = P(\Omega) = \sum_{\omega \in \Omega} P(\{\omega\}) = \sum_{\omega \in \Omega} p = p \sum_{\omega \in \Omega} 1 = p|\Omega|.$$

$$p = \frac{1}{|\Omega|},$$

*We see that for $A \subseteq \Omega$ as*

$$P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p = p \sum_{\omega \in A} 1 = p|A| = \frac{|A|}{|\Omega|}.$$

*Then for any $A \in \mathcal{F}$*

$$P(A) = \frac{|A|}{|\Omega|}$$

**Example 6.** $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$ *- a countable set. Let* $p_1, \dots, p_n, \dots$ *- sequence of non-negative numbers s.t.* $\sum_{k=1}^{\infty} p_k = 1$*. We can choose* $\mathcal{F} = 2^{\Omega}$ *and* $P(\{\omega_i\}) = p_i$*. This choice defines the probability space* $(\Omega, \mathcal{F}, P)$*, and for any* $A \in \mathcal{F}$ *we have*

$$P(A) = \sum_{k=1}^{\infty} \mathbf{1}_A(\omega_k) p_k$$

*where*

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

**Example 7.** ***Uniform probability****: We choose* $\Omega \subset \mathbb{R}^d$*, that is* $\Omega$ *is a Borel subset of* $\mathbb{R}^d$ *and we assume that* $0 < |\Omega| < \infty$*, where* $|\Omega| = \int_{\mathbb{R}^d} \mathbf{1}_\Omega$ *is the Lebesgue measure*[1]*. Let* $\mathcal{F} = \mathcal{B}(\Omega)$ *the smallest* $\sigma-$*algebra that contains all open subsets of* $\Omega$ *and for* $A \in \mathcal{F} = \mathcal{B}(\Omega)$*,*

$$P(A) = \frac{|A|}{|\Omega|}$$

*Then* $(\Omega, \mathcal{F}, P)$ *is a probability space. We use this probability space to describe experiments where a point(s) are randomly chosen from* $\Omega$*.*

[1]This integral is actually the Lebesgue integral. For a Riemann integrable function, the Lebesgue integral is equal to the Riemann integral. We will always consider subsets of $\mathbb{R}^d$ whose characteristic functions are Riemann integrable.

# Combinatorial

As we saw earlier, if the sample space is finite and we can assume equal probability, the problems of calculating probabilities are reduced to being able to compute the sizes of the sets involved. In other words, we need to be good at counting the number of elements in various sets. The science of counting is called combinatorics. Next, we will consider some simple combinatorial rules and their application in probability theory when a uniform distribution is appropriate.

## Counting Permutations

Suppose four friends go to a restaurant, and each checks his or her coat. At the end of the meal, the four coats are randomly returned to the four people. What is the probability that each of the four people gets his or her own coat? Here the total number of different ways the coats can be returned is equal to $4 \times 3 \times 2 \times 1$, or $4!$ (i.e., four factorial). This is because the first coat can be returned to any of the four friends, the second coat to any of the three remaining friends, and so on. Only one of these assignments is correct. Hence, the probability that each of the four people gets his or her own coat is equal to $\frac{1}{4!}$, or $\frac{1}{24}$.

Here we are counting permutations, or sequences of elements from a set where no element appears more than once. We can use the multiplication principle to count permutations more generally. For example, suppose that we have the set $S$ with $n$ different objects and we want to count the number of permutations of length $k \leq n$ obtained from $S$, i.e., we want to count the number of elements of the set

$$\{s_1, \ldots, s_k : s_i \in S, s_i \neq s_j \text{ when } i \neq j\}$$

Then we have $n$ choices for the first element $s_1$, $n-1$ choices for the second element, and finally $n-k+1$ choices for the last element. So there are

$$n(n-1)\ldots(n-k+1)$$

permutations of length $k$ from a set of $n$ elements. This can also be written as $\frac{n!}{(n-k)!}$. Notice that when $k = n$ there are $n! = n(n-1)\ldots 2 \cdot 1$.

## Counting Subsets

Suppose 10 fair coins are flipped. What is the probability that exactly seven of them are heads? Here each possible sequence of 10 heads or tails (e.g., H H H T T T H T T T, T H T T T T H H H T, etc.) is equally likely, and by the multiplication principle the total number of possible outcomes is equal to 2 multiplied by itself 10 times, or $2^{10} = 1024$. Hence, the probability of any particular sequence occurring is $\frac{1}{1024}$. But of these sequences, how many have exactly seven heads?

To answer this, notice that we may specify such a sequence by giving the positions of the seven heads, which involves choosing a subset of size 7 from the set of possible indices $\{1, \ldots, 10\}$. There are $\frac{10!}{3!} = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4$ different permutations of length 7 from $\{1, \ldots, 10\}$, and each such permutation specifies a sequence of seven heads and three tails. But we can permute the indices specifying where the heads go in 7! different ways without changing the sequence of heads and tails. So the total number of outcomes with exactly seven heads is equal to $\frac{10!}{3!7!} = 120$. The probability that exactly seven of the 10 coins are heads is therefore equal to $\frac{120}{1024}$, or just under 12%.

In general, if we have a set $S$ of $n$ elements, then the number of different subsets of size $k$ that we can construct by choosing elements from $S$ is

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

which is called the binomial coefficient. This follows by the same argument, namely, there are $\frac{n!}{(n-k)!}$ permutations of length $k$ obtained from the set; each such permutation, and the $k!$ permutations obtained by permuting it, specify a unique subset of $S$.

## Counting Sequences of Subsets and Partitions

When we want to divide a larger set into several smaller, non-overlapping subsets, we can use a powerful counting method based on the multiplication principle. For example, how many different ways can a deck of 52 cards be divided up into four hands of 13 cards each, with the hands labelled North, East, South, and West, respectively?

1. **North Hand ($N$):** Choose 13 cards from 52: $\binom{52}{13}$
2. **East Hand ($E$):** 39 cards remain. Choose 13 from 39: $\binom{39}{13}$
3. **South Hand ($S$):** 26 cards remain. Choose 13 from 26: $\binom{26}{13}$
4. **West Hand ($W$):** 13 cards remain. Choose 13 from 13: $\binom{13}{13} = 1$

The total number of ways is the product of these combinations:

$$\text{Total Ways} = \binom{52}{13} \cdot \binom{39}{13} \cdot \binom{26}{13} \cdot \binom{13}{13}$$

Expanding the factorial terms shows a cancellation pattern (telescoping product):

$$\text{Total Ways} = \left(\frac{52!}{13!39!}\right) \cdot \left(\frac{39!}{13!26!}\right) \cdot \left(\frac{26!}{13!13!}\right) \cdot 1 = \frac{52!}{13!13!13!13!}$$

This equals

$$\binom{52}{13, 13, 13, 13} = \frac{52!}{13!13!13!13!} \approx 5.364 \times 10^{28}$$

which is a very large number.

In general, suppose we have a set $S$ of $n$ elements and we want to count the number of elements of

$$\{S_1, S_2, \dots, S_l : S_i \subset S, |S_i| = k_i, S_i \cap S_j = \emptyset \text{ when } i \neq j\}$$

namely, we want to count the number of sequences of $l$ subsets of a set where no two subsets have any elements in common and the $i$-th subset has $k_i$ elements. By the multiplication principle, this equals

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-k_1-\cdots-k_{l-1}}{k_l} = \frac{n!}{k_1!\cdots k_l!(n-k_1-\cdots-k_l)!}$$

because we can choose the elements of $S_1$ in $\binom{n}{k_1}$ ways, choose the elements of $S_2$ in $\binom{n-k_1}{k_2}$ ways, etc.

When we have that $S = S_1 \cup S_2 \cup \cdots \cup S_l$, in addition to the individual sets being mutually disjoint, then we are counting the number of ordered partitions of a set of $n$ elements with $k_1$ elements in the first set, $k_2$ elements in the second set, etc. In this case, the previous expression equals

$$\binom{n}{k_1, k_2, \dots, k_l} = \frac{n!}{k_1!k_2!\cdots k_l!}$$

which is called the multinomial coefficient.