Probability and Random Processes

Malek Ebadi

May 2022

Contents

1 Probability: Intuition, Formalism **1**
- 1.1 Uncertainty, randomness, and probability . . . 1
- 1.2 Algebra of Sets . . . 2
- 1.3 Set Operations . . . 4
- 1.4 Venn Diagrams . . . 4
- 1.5 Probability of an event . . . 4
- 1.6 Axioms of Probability . . . 6

2 Combinatorics and Counting **9**
- 2.1 Introduction to Combinatorics and Counting Thechniques . . . 9
- 2.2 Additive Principle . . . 10
- 2.3 Multiplicative Principle . . . 11
- 2.4 Permutations . . . 12
- 2.5 Combinations . . . 18
  - 2.5.1 Binomial coefficients. . . . 20
  - 2.5.2 Pascal's Triangle . . . 21
  - 2.5.3 Observations From Pascal's Triangle . . . 22
- 2.6 Probabilities Involving Combinatorics . . . 24

3 Conditional Probability and Independence **27**
- 3.1 Conditional Probability . . . 27
- 3.2 Bayes' theorem . . . 28
  - 3.2.1 A side Note About Bayesian Inference . . . 34
- 3.3 Independence Of Events . . . 35

Chapter 1

Probability: Intuition, Formalism

1.1 Uncertainty, randomness, and probability

Frequentist Approach to Probability: Imagine the experiment of flipping a coin. This experiment is known to be a random experiment. (I admit, there exist serious controversies). It is random, because there is uncertainty beforehand about which side of the coin will appear when the coin is flipped. But what does it exactly mean, when we say *the probability of the next flip being heads is 0.5?*

When faced with an uncertain event, the frequentist approach tends to repeat the experiment over and over again. According to this approach probability of an event represents the *relative frequency* with which an event occurs after conducting the same experiment infinitely many times. Since conducting an experiment infinitely many times is not possible, the results in the limit are considered, meaning that for instance in coin flip experiment, the experiment is repeated until the observed frequency is a good estimate of the true probability of occurrence of a particular event.

Bayesian Approach to Probability: In many cases probabilities can not be associated with repeated trials. For instance, the probability that it will rain in a certain day. In such settings Bayesian approach introduces subjective estimates of probabilities, representing the *degree of belief* that an event occurs. This approach is interrelated with Bayesian statistics, a field of statistics that deals with modeling uncertainty and has useful applications in many fields including machine learning and artificial intelligence.

1.2 Algebra of Sets

Definition 1. *(Outcome)*
An outcome ω *of a random experiment is one possible realization of conducting the experiment.*

- Flipping a coin: $\omega_1 = \text{H}$, $\omega_2 = \text{T}$

Definition 2. *(Sample space)*
A sample space Ω *is the set of all possible outcomes of an experiment.*

1. Flipping a coin: $\Omega = \{\text{H}, \text{T}\}$
2. Rolling a dice experiment: $\Omega = \{1, 2, 3, 4, 5, 6\}$
3. Three rounds of Russian roulette experiment: $\Omega = \{\text{D}, \text{LD}, \text{LLD}\}$.

Definition 3. *(Event)*
An event is a set of outcomes contained in the sample space, Ω. An event is a subset of Ω.

1. Flipping a coin: $\Omega = \{\text{H}, \text{T}\}$
   - Event A: getting H $\Rightarrow \text{A} = \{\text{H}\} \in \Omega$
2. Flipping two coins: $\Omega = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$
   - Event A: getting H,H $\Rightarrow \text{A} = \{\text{H}, \text{H}\} \in \Omega$
   - Event B: getting T,H $\Rightarrow \text{B} = \{\text{T}, \text{H}\} \in \Omega$
   - Event C: getting H,T $\Rightarrow \text{C} = \{\text{H}, \text{T}\} \in \Omega$

Definition 4. *(algebra)*
A set of events, $\mathcal{F}$, is an algebra if

i: $\text{A} \in \mathcal{F}$ *implies that* $\text{A}^{\text{c}} \in \mathcal{F}$.
ii: $\text{A} \in \mathcal{F}$ *and* $\text{B} \in \mathcal{F}$ *implies that* $\text{A} \cup \text{B} \in \mathcal{F}$ *and* $\text{A} \cap \text{B} \in \mathcal{F}$.
iii: $\Omega \in \mathcal{F}$ *and* $\emptyset \in \mathcal{F}$.

The events in $\mathcal{F}$ are called *measurable events in $\mathcal{F}$*.

Definition 5. *(σ-algebra)*
A non-empty collection $\mathcal{A}$ of subsets of a set Ω is called a σ-algebra if given $\text{A}, \text{A}_1, \text{A}_2, \dots \in \mathcal{A}$ *we have*

i: $\text{A}^{\text{c}} \in \mathcal{A}$
ii: $\bigcup_n \text{A}_\text{n} \in \mathcal{A}$

iii: $\bigcap_n \mathrm{A_n} \in \mathcal{A}$

The definition implies that an algebra of sets ($\mathcal{F}$) is a $\sigma-$algebra ($\mathcal{A}$), if it is *closed under countable intersections*. Suppose $\mathrm{A_n} \in \mathcal{F}$ is a countable family of events measurable in $\mathcal{F}$, and $\mathrm{A} = \cap_\mathrm{n} \mathrm{A_n}$ is the set of outcomes in all of the $\mathrm{A_n}$, then $\mathrm{A} \in \mathcal{F}$, too.

Proposition 1. *An algebra closed under countable intersections is also closed under countable unions, and conversely.*

An algebra is automatically a $\sigma-$algebra if Ω is finite. If Ω is infinite, an algebra might or might not be a $\sigma-$algebra.

Example 1.1. *Let Ω be the set of integers and $\mathrm{A} \in \mathcal{F}$ if A is finite or $\mathrm{A^c}$ is finite. This $\mathcal{F}$ is an algebra (check), but not a $\sigma-$algebra. For example, if $\mathrm{A_n}$ leaves out only the first n odd integers, then A is the set of even integers, and neither A nor $\mathrm{A^c}$ is finite.*

□

In a $\sigma-$algebra, it is possible to take limits of infinite sequences of events, just as it is possible to take limits of sequences of real numbers.

Lemma 1. *Let σ-algebra $\mathcal{A}$ in Ω, and $\mathrm{A_1}, \mathrm{A_2}, \cdots \in \mathcal{A}$,*

i: $\Omega \in \mathcal{A}$
ii: $\emptyset \in \mathcal{A}$

Proof. Since $\mathcal{A}$ is non empty, we can find $\mathrm{A} \in \mathcal{A}$. Thus $\Omega = \mathrm{A} \bigcup \mathrm{A^c} \in \mathcal{A}$. Then taking complements shows $\emptyset \in \mathcal{A}$. □

Lemma 2. *Given a class $\mathcal{C}$ of σ-algebras on Ω, the intersection is also a σ-algebra.*

Proof. Because we have shown that every σ-algebra contains Ω, we know that the intersection is non-empty. Now let $\mathrm{A}, \mathrm{A_1}, \mathrm{A_2}, \ldots$ be in every σ-algebra. Clearly every σ-algebra in the class contains $\bigcap_n \mathrm{A_n}$, hence so does the intersection. Similarly with $\bigcup_n \mathrm{A_n}$ and $\mathrm{A^c}$. □

Note that a union of σ-algebras is not necessarily a σ-algebra. However, a union of σ-algebras generates a σ-algebra in an appropriate sense.

Definition 6. *Given a collection $\mathcal{C}$ of subsets of Ω, we let $\sigma(\mathcal{C})$ be the smallest σ-algebra containing $\mathcal{C}$.*

Note that the definition makes sense since the set of all subsets of Ω is a σ-algebra. Therefore, the class of σ-algbras containing $\mathcal{C}$ is non-empty and $\sigma(\mathcal{C})$ is the intersection of of the class by the previous lemma.

1.3 Set Operations

As discussed, events are set of outcomes. Hence, set operations also apply to them.

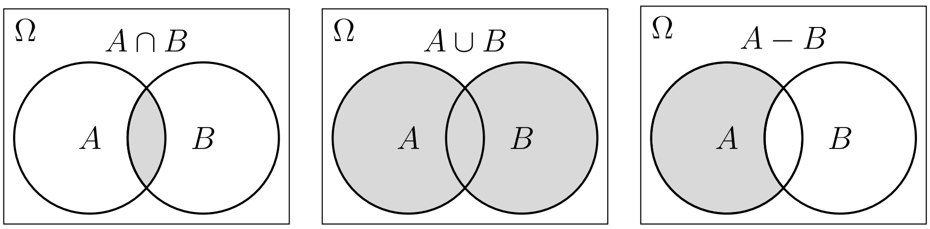
Definition 7. Union, Intersection and Complement of Events

1. *The* **union** *of two events* A and B denoted $A \cup B$ consists of outcomes that are either in A or B.
2. The **intersection** of two events A and B denoted $A \cap B$ consists of outcomes that both in A and B.
3. The **complement** of an event A denoted A^c is the set of all outcomes in Ω not contained in A, $A^c = \Omega \backslash A$.

Definition 8. *Two events* A and B are **mutually exclusive** or **disjoint** if they have no outcomes in common, i.e. $A \cap B = \emptyset$.

1.4 Venn Diagrams

A Venn diagram is a graphical representation of sets and set operations. Each Venn diagram includes a rectangle representing the universal set, and circle(s) inside the rectangle representing sets. Figure 1.1 shows a few examples for two sets.



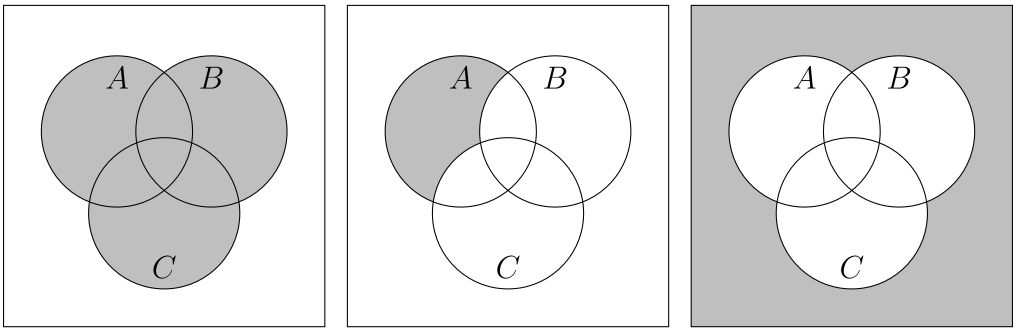
(a) Intersection (b) Union (c) Subtraction

Figure 1.1: Venn diagrams for two sets

Figure 1.2 shows examples of Venn diagrams with three sets.

1.5 Probability of an event

Given an experiment and a sample space Ω, the objective of probability is to assign to each event A a number $\mathbb{P}(A)$, called the probability of the event



(a) $(A \cup B \cup C)$ (b) $A - (B \cup C)$ (c) $(A \cup B \cup C)^c$

Figure 1.2: Venn diagrams

A, which will give a precise measure of the chance that A will occur. In an experiment with N outcomes that are equally likely the probability of any outcome A is $\mathbb{P}(\mathrm{A}) = 1/\mathrm{N}$.

Definition 9. *The probability of an event is the sum of the probabilities of the outcomes that make up the event*

$$\mathbb{P}(\mathrm{A}) = \sum_{\omega \in \mathrm{A}} \mathrm{p}(\omega) \ .$$

Example 1.2. *Suppose we toss a coin 4 times, at each time either a H (heads) or T (tails) appears. The outcomes are:* TTTT, TTTH, TTHT, TTHH, THTT, ..., HHHH. Hence, the sample space is

$$\Omega = \{\mathrm{TTTT}, \mathrm{TTTH}, \mathrm{TTHT}, \mathrm{TTHH}, \mathrm{THTT}, \ldots, \mathrm{HHHH}\}.$$

The number of outcomes in the sample space is 16, that is $|\Omega| = 16$, where $| \,.\, |$ represents the cardinality of a set. If each outcome ω is equally likely, then,

$$p(\omega) = \frac{1}{16} \quad \forall \ \omega \in \Omega.$$

Let A be the event that the first two tosses are H, then A includes 4 outcomes,

$$\mathrm{A} = \{\mathrm{HHHH, \ HHHT, \ HHTH, \ HHTT}\} \, ,$$

each having probability 1/16. Therefore,

$$\text{Prob}\{\text{first two are H}\} = \mathbb{P}(\mathrm{A}) = \sum_{\omega \in \mathrm{A}} \mathrm{p}(\omega) = \sum_{\omega \in \mathrm{A}} \frac{1}{16} = \frac{4}{16} = \frac{1}{4} \ .$$

□

1.6 Axioms of Probability

Axiom 1. *For any event* B, $0 \leq \mathbb{P}(\mathrm{A}) \leq 1$.

Axiom 2. $\mathbb{P}(\Omega) = 1$, *and therefore* $\mathbb{P}(\emptyset) = 0$.

Axiom 3. *If* $\mathrm{A_1}, \mathrm{A_2}, \ldots, \mathrm{A_n}$ *are mutually exclusive events, then*

$$\mathbb{P}\left(\mathrm{A_1} \cup \mathrm{A_2} \cup \cdots \cup \mathrm{A_n}\right) = \sum_{i=1}^{n} \mathbb{P}(\mathrm{A_i}).$$

Axiom 3 can be generalized to infinitely many mutually exclusive events when $n \to \infty$,

$$\mathbb{P}\left(\mathrm{A_1} \cup \mathrm{A_2} \cup \cdots\right) = \sum_{i=1}^{\infty} \mathbb{P}(\mathrm{A_i}).$$

In the experiment of flipping a coin $\Omega = \{\mathrm{H}, \mathrm{T}\}$ and also $\mathrm{H} \cup \mathrm{T} = \Omega$.

$$\mathbb{P}(\mathrm{H} \cup \mathrm{T}) = \mathbb{P}(\mathrm{H}) + \mathbb{P}(\mathrm{T}) = \mathbb{P}(\Omega) = 1$$

Proposition 2. *For any event* A, $\mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{A^c}) = 1$ or, $\mathbb{P}(\mathrm{A}) = 1 - \mathbb{P}(\mathrm{A^c})$.

In the experiment of flipping a coin $\mathbb{P}(\mathrm{T}) = 1 - \mathbb{P}(\mathrm{H})$, since one event is the complement of the other.

Example 1.3. *Suppose we flip an unfair coin n times for which* $\mathbb{P}(\mathrm{H}) = \mathrm{p}$. *We are interested in the probability of the event that we observe more than one heads?*
Let A be the event that we observe more than one heads. Notice that $\mathrm{A^c}$ will be the event that zero heads are observed.

$$\begin{aligned} \mathbb{P}(\mathrm{A^c}) &= \text{Prob}\{\text{all flips land on tail}\} = (1-p)^n \\ \mathbb{P}(\mathrm{A}) &= 1 - (1-p)^n. \end{aligned}$$

□

Proposition 3. *If* A and B are mutually exclusive, then $\mathbb{P}\left(\mathrm{A} \cap \mathrm{B}\right) = 0$ and $\mathrm{A} \cap \mathrm{B} = \emptyset$.

Proposition 4. *For any two events* A and B,

$$\mathbb{P}\left(\mathrm{A} \cup \mathrm{B}\right) = \mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{B}) - \mathbb{P}\left(\mathrm{A} \cap \mathrm{B}\right).$$

For any three events A, B, C,

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ &\quad - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) \\ &\quad + \mathbb{P}(A \cap B \cap C).\end{aligned}$$

Chapter 2

Combinatorics and Counting

2.1 Introduction to Combinatorics and Counting Thechniques

In discrete mathematics, combinatorics refer to techniques used for counting. Some of the basic principles used in combinatorics include additive and multiplicative principles, permutations, combinations, binomial coefficients and Pascal's triangle.

Definition 10 (Some definitions related to counting)**.**

Distinct Objects: *all of the objects that we want to arrange or select are distinct.*
Example: Letters in $CHEMISTRY$*.*

Identical Objects: *some of the objects that we want to arrange or select are identical.*
Example: Letters in $STATISTICS$ *are not all distinct,* $STATISTICS$*.*

Order is important: *when the order matters,* CAB *is different than* ABC*.*

Order is not important: *When the order does not matters, CAB is same as ABC.*

Sampling without replacement: *means that after choosing an item, it is thrown away and cannot be chosen again. In this case, we are not allowed to choose one object repeatedly.*
Example: If we want to form three letter words from ABC, we are not allowed to form BBB.

Sampling with replacement: *means that after choosing at item, it is put back and can be chosen again. We are allowed to choose one object repeatedly.*

2.2 Additive Principle

Disjoint Sets. The *additive principle* states that for two disjoint sets A and B, the cardinality of their union is the sum of their cardinalities,

$$|A \cup B| = |A| + |B|.$$

The additive principle can be generalized to n sets. Given n pairwise disjoint sets $A_1, A_2, \ldots, A_n$, then

$$|A_1 \cup \cdots \cup A_n| = |A_1| + \cdots + |A_n|.$$

When sets are not disjoint. The inclusion-exclusion principle generalizes the additive principle to when the sets aren't disjoint.

Definition 11 (Inclusion-exclusion principle)**.** *For two sets A and B,*

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

For three sets A, B, and C

$$|A \cup B \cup C| = |A| + |B| + |C| \\ - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

2.3 Multiplicative Principle

The multiplicative principle is about making choices in stages. In the simplest form of it, the multiplicative principle states that if you have m choices, and for each choice you have n second choices, then altogether you have mn choices.

Example 2.1. *Suppose you want to choose an optional course from your curriculum and you have two options: French or Psychology. Both Mr. Jones and Mr. Smith offer French course. Psychology course is offered by three instructors: Mr. Brown, Mr. Frenk and Mr. Boyd. Then, in total you have* $2 \times 3 = 6$ *choices.*

□

Remark 1. *The cardinality of Cartesian product of two sets A and B, denoted as* $|\mathrm{A} \times \mathrm{B}|$ *also follows from the multiplicative principle.*

$$|\mathrm{A} \times \mathrm{B}| = |\mathrm{A}|\,|\mathrm{B}|.$$

Remember that the Cartesian product $\mathrm{A} \times \mathrm{B}$ *consists of all ordered pairs* (a, b) *such that* $a \in \mathrm{A}$ *and* $b \in \mathrm{B}$*. Formally,*

$$\mathrm{A} \times \mathrm{B} = \{(\mathrm{a}, \mathrm{b}) : \mathrm{a} \in \mathrm{A}, \mathrm{b} \in \mathrm{B}\}$$

Example 2.2. *Suppose set* $\mathrm{A} = \{0, 1, 2\}$ *and set* $\mathrm{B} = \{3, 4\}$*. Then the Cartesian product* $\mathrm{A} \times \mathrm{B}$ *consists of 6 elements, since* $|\mathrm{A}| = 3$ *and* $|\mathrm{B}| = 2$*.*

$$\mathrm{A} \times \mathrm{B} = \{(0, 3), (0, 4), (1, 3), (1, 4), (2, 3), (2, 4)\}.$$

Proposition 5 (General Form of Multiplicative Principle)**.** *A set consists of ordered collections of k-tuples with n_1 choices for element 1, n_2 choices for element 2 ,..., and n_k choices for element k. The number of possible k-tuples are $n_1 n_2 \cdots n_k$.*

Remark 2. *Multiplicative principle applies to the product of sets. Given finite sets $A_1, A_2, \ldots, A_n$, their product $A_1 \times A_2 \times \cdots \times A_n$ consists of ordered n-tuples $(a_1, a_2, \ldots, a_n)$ where each a_i belongs to the corresponding set A_i. In order to choose one of these ordered n-tuples, for the first stage you have a choice of choosing any one of the elements of A_1 to be a_1. The number of choices at stage 1 is the cardinality of A_1. For the second stage you have*

$|A_2|$ choices for a_2, and so forth. The multiplicative principle gives us the standard formula for the cardinality of the product

$$|A_1 \times A_2 \times \cdots \times A_n| = \prod_{i=1}^{n} |A_i|.$$

Example 2.3 (Fixed price menu)**.** *A fixed price menu consists of the followings: appetizer, main food, and dessert. There are 3 types of appetizers, 4 types of main dish, and 2 types of desserts.*
Appetizers: $n_1 = 3$, Mains: $n_2 = 4$, Desserts $n_3 = 2$. Hence, the number of meals is: $n_1 \times n_2 \times n_3 = 24$. Let a_i, m_i and d_i denote type i of each item in the menu. Table 2.1 lists all possible menus.

(a_1, m_1, d_1)	(a_1, m_1, d_2)	(a_1, m_2, d_1)	(a_1, m_2, d_2)
(a_1, m_3, d_1)	(a_1, m_3, d_2)	(a_1, m_4, d_1)	(a_1, m_4, d_2)
(a_2, m_1, d_1)	(a_2, m_1, d_2)	(a_2, m_2, d_1)	(a_2, m_2, d_2)
(a_2, m_3, d_1)	(a_2, m_3, d_2)	(a_2, m_4, d_1)	(a_2, m_4, d_2)
(a_3, m_1, d_1)	(a_3, m_1, d_2)	(a_3, m_2, d_1)	(a_3, m_2, d_2)
(a_3, m_3, d_1)	(a_3, m_3, d_2)	(a_3, m_4, d_1)	(a_3, m_4, d_2)

Table 2.1: List of all possible menus.

□

2.4 Permutations

Permutation is about arrangement. A permutation is an arrangement of objects without repetition when order is important. Depending on whether objects are identical or distinct and whether we arrange all of the objects or only some of them, we will use different counting techniques.

Permutation of All Distinct Objects. Suppose we have a set A of n elements. How may permutations are there on set A?

Proposition 6. *(Permutation using all of the distinct objects)*
A permutation of n distinct objects, arranged into one group of size n, without

repetition, and order being important is:

$$nPn = P(n,n) = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$$

Obviously, in the first stage, we have a choice among one of the n elements to go first. In the second stage, there are $n-1$ remaining elements, and choose one of them to go second. At the next stage, chose one of the remaining $n-2$ elements to go next. And so forth until the last stage, when there's only one element left, so it goes last. Thus, the number of permutations of a set of n elements is

$$n(n-1)(n-2)\cdots 2 \cdot 1.$$

This last expression is usually abbreviated $n!$ and read "n factorial" or "factorial n".

Example 2.4. *Find all permutations of the letters in set $A = \{a, b, c\}$.*

Since we are using all three objects, we can arrange the letters in $3P3 = P(3,3) = 3! = 6$ *ways, which are*

$$abc, acb, bac, bca, cab, cba.$$

When choosing a permutation of "a,b,c", there are 3 choices for the first letter, 2 remaining choices for the second (since we cannot choose the first letter again), 1 remaining choice for the third. Hence, there are $3 \cdot 2 \cdot 1 = 6$ *choices altogether.*

□

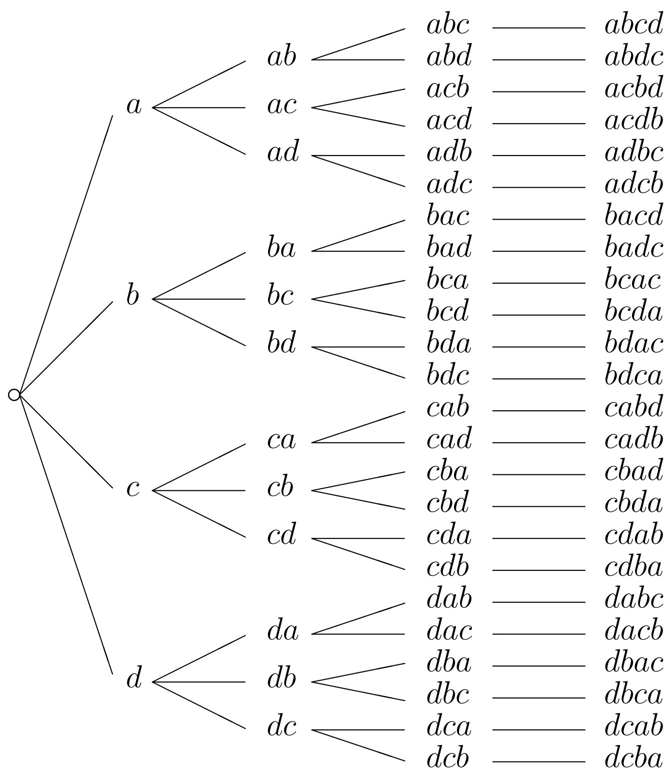
Example 2.5. *Find all permutations of the letters in set $A = \{a, b, c, d\}$.*

Since we are using all four objects, we can arrange the letters in $4! = 24$ *ways, with 4 choices for the first letter, 3 choices for the second, 2 for the third letter and 1 for the last letter,* $4 \times 3 \times 2 \times 1 = 4!$.

abcd	*bacd*	*cabd*	*dabc*
abdc	*badc*	*cadb*	*dacb*
acbd	*bcad*	*cbad*	*dbac*
acdb	*bcda*	*cbda*	*dbca*
adbc	*bdac*	*cdab*	*dcab*
adcb	*bdca*	*cdba*	*dcba*

□

Tree diagrams. *Tree diagram* can be used to illustrate permutations. Consider Example 2.5, where we choose a permutation of the 4 letters from a set $A = \{a, b, c, d\}$. The choices will be made in 4 stages according to the following diagram:



The first stage chooses one of the four letters to go first. That gives us our first branching of the tree at the left. After we've taken that branch, we'll be at one of the four *nodes* or *states* labelled a, b, c, or d. At this second stage, we choose a second letter that can't be the same as the first. In each case we have three choices this time, so we'll take one of the three branches to get to a state labelled by two letters. At the third stage, we've got two choices, so for each state there are two branches leading to a state labelled with three letters. At this state the last letter is determined, so there's only one branch to a *leaf* of the tree.

Sterling's approximation for factorials. The factorial function $n!$ grows very fast with n. James Sterling (1692–1770) proposed an approximation method to compute factorials of large numbers. Named after James Sterling, the Sterling's approximation is

$$n! \approx n^n e^{-n} \sqrt{2\pi n}$$

This approximation is fairly good even for numbers as small as 10 where the approximation has an error of less than 1%. It's accuracy increases with n.

n	$n!$	approx	ratio
1	1	0.922137	1.084
2	2	1.91900	1.042
3	6	5.83621	1.028
4	24	23.5062	1.021
5	120	118.019	1.016
6	720	710.078	1.014
7	5040	4980.40	1.012
8	40320	39902.4	1.011
9	362880	359536	1.0093
10	3628800	3598690	1.0084
11	39916800	39615600	1.0076
12	479001600	475687000	1.0070

Permutation of k Distinct Objects from n Objects. One variant of permutations is the case when we don't want complete permutations of a set of n elements, rather we are interested in partial permutations, say k-*permutation.* If $k \leq n$, a k*-permutation* is an ordered listing of just k elements of a set of n elements.

Proposition 7. *(Permutation using some of the distinct objects)*
A permutation of n (distinct) objects, arranged in groups of size k, without repetition, and order being important is:

$$nPk = P(n,k) = \frac{n!}{(n-k)!}$$

We can use the multiplicative principle to determine the k-permutations of a set of n elements. In the first stage, we have a choice among n elements to go first. In the second stage, there are $n-1$ remaining elements to go second. At the third stage, $n-2$ elements to go third. Continuing in this fashion, at the kth stage, there will be $(n-k)+1$ remaining elements. Hence, the number of k-permutations of a set of n elements is

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

where the equality follows from the fact that

$$[n(n-1)(n-2)\cdots(n-k+1)] \times (n-k)! = n!$$

Example 2.6. *Find all two-letter permutations of the letters in set* $A = \{a, b, c\}$.
Since we are using two of three objects, we can arrange the letters in $3P2 = P(3, 2) = 3! = 6$ *ways, which are*

$$ab, ac, ba, bc, ca, cb.$$

□

Example 2.7. *Find all two-letter permutations of the letters in set* $A = \{a, b, c, d\}$.

ab	*ba*	*ca*	*da*
ac	*bc*	*cb*	*db*
ad	*bd*	*cd*	*dc*

□

Example 2.8. *Find all three-letter permutations of the letters in set* $A = \{a, b, c, d\}$.

abc	*bac*	*cab*	*dab*
abd	*bad*	*cad*	*dac*
acb	*bca*	*cba*	*dba*
acd	*bcd*	*cbd*	*dbc*
adb	*bda*	*cda*	*dca*
adc	*bdc*	*cdb*	*dcb*

□

Permutation When There Are Identical Objects. Another variant of permutations is the case when the elements in the set are identical. For instance, when we want to count the permutations of letters in set $A = \{S, T, A, T, I, S, T, I, C, S\}$. Obviously there are repetitions in letters, so exchanging them will not create different permutation.

Proposition 8. *(Permutation using all of the objects some of which are identical)*
A permutation of n objects such that n_1 of them are identical of type 1, n_2 of them are identical of type 2, ..., and n_k of them are identical of type k, (i.e. $n = n_1 + n_2 + \cdots + n_k$), without repetition, and order being important is:

$$P\big[n, (n_1, n_2, \ldots, n_k)\big] = \frac{n!}{n_1! \times n_2! \times \ldots n_k!}$$

Example 2.9. *Find all permutations of the letters in set $A = \{B, O, B\}$.*

Remind that without repetition means we are not allowed to form "OOO", but we are allowed to form "BOB", as if first B is different than second B. Order matters means that "BOB" is different than "OBB".

One way to approach this question is to first assume that letters are all distinct. Let's write the second "B" as "b". We already know that arranging 3 objects out of 3 (without repetition, and order being important) can be done in $nPn = 3P3 = 3! = 6$ ways, which are BOb, BbO, OBb, ObB, bBO, bOB. However, if we write "b"'s in these 6 arrangements as "B", we will get BOB, BBO, OBB, OBB, BBO, BOB. In fact, there are two repetitions from each word BOB, BBO, OBB. To what number should we divide 6 to get rid of repetitions? The answer is $2!$. In $2!$ ways we are repeating these three words. So the final answe will be $6/2! = 3$ ways which are BOB, BBO, OBB.

□

If you are still not convinced, consider the following example which contains 3 identical letter.

Example 2.10. *Consider the word "LULL". Let's imagine this word as $L_1UL_2L_3$. There are $4! = 24$ ways to permute them.*

$UL_1L_2L_3$	$UL_1L_3L_2$	$UL_2L_1L_3$	$UL_2L_3L_1$	$UL_3L_1L_2$	$UL_3L_2L_1$
$L_1UL_2L_3$	$L_1UL_3L_2$	$L_1L_2UL_3$	$L_1L_2L_3U$	$L_1L_3UL_2$	$L_1L_3L_2U$

$L_2UL_1L_3$	$L_2UL_3L_1$	$L_2L_1UL_3$	$L_2L_1L_3U$	$L_2L_3UL_1$	$L_2L_3L_1U$
$L_3UL_1L_2$	$L_3UL_2L_1$	$L_3L_1UL_2$	$L_3L_1L_2U$	$L_3L_2UL_1$	$L_3L_2L_1U$

Now let's revert them back to simple "l".

$ULLL$	$ULLL$	$ULLL$	$ULLL$	$ULLL$	$ULLL$
$LULL$	$LULL$	$LLUL$	$LLLU$	$LLUL$	$LLLU$
$LULL$	$LULL$	$LLUL$	$LLLU$	$LLUL$	$LLLU$
$LULL$	$LULL$	$LLUL$	$LLLU$	$LLUL$	$LLLU$

We observe that there are 6 "ULLL", 6 "LULL", 6 "LLUL" and 6 "LLLU". $6 \times (\textit{number of unique words}) = 24$, *with* $6 = 3!$ *and* $24 = 4!$.

□

Example 2.11. *Consider the word "STATISTICS". Here are the frequency of each letter:* $S = 3, T = 3, A = 1, I = 2, C = 1$, *there are 10 letters in total.*

$$P\big[10, (3, 3, 1, 2)\big] = \frac{10!}{3! \times 3! \times 1! \times 2!} = 50400$$

□

2.5 Combinations

Combination is about selection. A combination is a selection of objects without repetition where order is not important. Combination is also called "Committee Selection". Imagine a committee to be selected for the thesis defense of a Ph.D. student. It is not possible to select same person twice for the committee. It makes also no difference whether we select person "A" before "B", since a committee made of "A,B" (A selected before B) is the same as a committee made of "B,A" (B selected before A).

Notice that in neither permutation nor combination repetition is not allowed. Hence, the only difference between the definition of a permutation and a combination is whether order is important or not. Textbooks may use

ordered and **u**nordered sets to refer **o**rder being important and **o**rder being not important respectively.

Proposition 9. *A combination of n distinct objects, arranged in groups of size k, without repetition, and order being important is:*

$$nCr = C(n,k) = \binom{n}{k} = \frac{n!}{(n-k)! \times k!}$$

A *combination* of size k from a set S of size n is just a subset of size k. It's more often it's called a k-subset when the size is specified. A k-subset is related to k-permutations but they're not the same. A k-permutation is a listing of k distinct elements of S where the order of the elements in the listing is relevant. But for a k-subset, the elements are not listed in any particular order; that is, order doesn't matter.

Let's take an example. Let S be the 5-element set $S = \{a, b, c, d, e\}$. There are $5 \cdot 4 \cdot 3 = 60$ 3-permutations of S, but there are far fewer 3-subsets of S. For instance, one 3-subset is $\{a, b, c\}$. But this subset is associated to 6 of the 3-permutations, namely, *abc*, *acb*, *bac*, *bca*, *cab*, and *cba*. There are 6, of course, because there are $3! = 6$ full permutations of a set of 3 elements.

In general, each k-subset is associated to $k!$ of the k-permutations. Since there are $\frac{n!}{(n-k)!}$ of the k-permutations altogether, that implies that the number of k-subsets of a set of n elements is exactly $\frac{n!}{k!(n-k)!}$.

Example 2.12. *Find all two-letter combinations of the letters "ABC". Note that the question can be framed as "how many committees can be made from "A,B,C". Knowing that AB = BA, AC = CA, BC = CB, there are only three committees.*

$$3C2 = C(3,2) = \binom{3}{2} = \frac{3!}{(3-2)! \times 2!} = 3$$

□

2.5.1 Binomial coefficients.

Combinations are used in the binomial theorem to give the coefficients of the expansion $(a+b)^n$. This fact is formalized in *Binomial Theorem.* Remember that binomials are polynomials with two terms.

Theorem 1. *(Binomial Theorem).*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} x^k y^{n-k}.$$

for $0 \leq k \leq n$ and $n \geq 0$, where $0! = 1$. The expression

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

is called a *binomial coefficient.* When $y = 1$, the binomial theorem reduces to:

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k.$$

One might wonder how combinations $\binom{n}{k}$ appear as the binomial coefficients. Let us first consider the expansion of binomial theorem when $n = 4$.

$$\begin{aligned}(x+y)^4 &= \binom{4}{0} x^0 y^4 + \binom{4}{1} x^1 y^3 + \binom{4}{2} x^2 y^2 + \binom{4}{3} x^3 y^1 + \binom{4}{4} x^4 y^0 \\ &= y^4 + 4x^3 y + 6x^2 y^2 + 4xy^3 + x^4\end{aligned}$$

Notice the coefficients when $n = 4$,

$$1 \quad 4 \quad 6 \quad 4 \quad 1$$

Consider the coefficient 6 of x^2y^2. There are 4 factors $(x+y)$ when $n = 4$.

$$(x+y)^4 = (x+y)(x+y)(x+y)(x+y)$$

When you expand the product you'll get a term x^2y^2 if you choose an x from exactly 2 of the 4 factors $(x+y)$, and y^2 coming from the remaining two factors. There are $\binom{4}{2} = 6$ ways of choosing 2 of the four factors, and each one contributes one x^2y^2, so the coefficient of x^2y^2 in the product will be $\binom{4}{2} = 6$.

2.5.2 Pascal's Triangle

Let's explore a few more cases from binomial coefficients. Consider the binomial $x + y$, with different powers n. The following illustrates $(x + y)^n$ for $n = 1, 2, 3, 4, 5$.

$$\begin{array}{rc}
(x+y)^0 = & 1 \\
(x+y)^1 = & x+y \\
(x+y)^2 = & x^2+2xy+y^2 \\
(x+y)^3 = & x^3+3x^2y+3xy^2+y^3 \\
(x+y)^4 = & x^4+4x^3y+6x^2y^2+4xy^3+y^4 \\
(x+y)^5 = & x^5{+}5x^4y{+}10x^3y^2{+}10x^2y^3{+}5xy^4{+}y^5
\end{array}$$

Let's extract the coefficients from each binomial.

$$\begin{array}{lccccccccccc}
n=0 & & & & & & 1 & & & & & \\
n=1 & & & & & 1 & & 1 & & & & \\
n=2 & & & & 1 & & 2 & & 1 & & & \\
n=3 & & & 1 & & 3 & & 3 & & 1 & & \\
n=4 & & 1 & & 4 & & 6 & & 4 & & 1 & \\
n=5 & 1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}$$

As discussed previously, the coefficients in these polynomials are combinations.

$$\begin{array}{ccccccccc}
 & & & & \binom{0}{0}=1 & & & & \\
 & & & / & & \backslash & & & \\
 & & & \binom{1}{0}=1 & & \binom{1}{1}=1 & & & \\
 & & / & \backslash & & / & \backslash & & \\
 & & \binom{2}{0}=1 & & \binom{2}{1}=2 & & \binom{2}{2}=1 & & \\
 & / & \backslash & / & & \backslash & / & \backslash & \\
 & \binom{3}{0}=1 & & \binom{3}{1}=3 & & \binom{3}{2}=3 & & \binom{3}{3}=1 & \\
/ & \backslash & / & \backslash & / & \backslash & / & \backslash & \\
\binom{4}{0}=1 & & \binom{4}{1}=4 & & \binom{4}{2}=6 & & \binom{4}{3}=4 & & \binom{4}{4}=1
\end{array}$$

Blaise Pascal (1623–1662) and Pierre de Fermat (1601–1665) studied these binomial coefficients in the context of probability in the 1600s. Their correspondence resulted in some of the first significant theory of probability and a systematic study of binomial coefficients. Because of Pascal's publication of their results, this particular arrangement of the binomial coefficients in a triangle is called *Pascal's triangle.*

2.5.3 Observations From Pascal's Triangle

There are lots of interrelations among these entries in Pascal's triangle.

Observation 1. Each element in the table is the sum of the two elements directly above it. The numbers along the sides are all 1s, and each entry in the middle is the sum of the two entries above it.

Observation 2. The sum of the numbers on a diagonal of Pascal's triangle equals the number below the last summand. For example, $1+2 = 3$, $1+2+3 = 6$, $1+3 = 4$, $1+3+6 = 10$, etc.

This fact is expressed formally in the identity:

$$\sum_{k=0}^{p} \binom{k+n-1}{n-1} = \binom{p+n}{n}$$

Here is one simple way to prove the identity. First observe that $\binom{k+n-1}{n-1}$ is the number of ways of dividing k objects into n subsets: line up $k+n-1$ objects and select $n-1$ of them to mark the boundaries of the n subsets. The number of ways of choosing $n-1$ from $k+n-1$ is, of course, $\binom{k+n-1}{n-1}$. Now the sum on the left hand side is the number of ways of dividing less than or equal to p objects into n subsets, one term for each number of objects $k = 0, \ldots, p$. The right hand side is the number of ways of dividing p objects into $n+1$ subsets. By ignoring the first subest, every way of dividing p objects into n subsets gives exactly one way of dividing k objects into $n-1$ subsets where k is p minus the number of elements in the ignored subset. Conversely, every way of dividing $k \leq p$ objects into n subsets gives rise to exactly one way of dividing p objects into $n+1$ subsets: just add a subset with $p-k$ objects. Therefore, the two sides of the above identity must be equal.

Observation 3. One important identity of the many important identities that hold for binomial coefficients is this one:

$$\binom{n}{k} = \binom{n}{n-k}$$

You can see why that's true in three different ways. First, they're both equal to $\dfrac{n!}{k!(n-k)!}$. Second as coefficients in the expansion of $(x+y)^n$, the coefficient of $x^k y^{n-k}$ is equal to the coefficient of $y^k x^{n-k}$. And third, each subset of k elements in a set of size n has a complement that has $n-k$ elements.

Observation 4. the recurrence relation

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1},$$

Observation 5. Note that in each row, n is fixed. Let's call that the n^{th} row; the top row is then the 0^{th} row. Note that the numbers in the n^{th} row sum to 2^n.

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n-1} + \binom{n}{1} = 2^n$$

For instance, when $n = 5$, we have $1 + 5 + 10 + 10 + 5 + 1 = 32 = 2^5$. That's because these binomial coefficients tell us the number of subsets of various sizes of a set of n elements. Since there are 2^n subsets in all, they have to add up to 2^n.

These binomial coefficients tell us the number of subsets of various sizes of a set of n elements. Since there are 2^n subsets in all, they have to add up to 2^n.

Proof. Using binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

and letting both x and y be 1, we get

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

□

2.6 Probabilities Involving Combinatorics

Example 2.13. *An urn contains 6 white and 9 black balls. If 4 balls are to be randomly selected without replacement, what is the probability that the first 2 selected are white and the last 2 selected are black?*
We define the events
W_1*: the first ball drawn is white,* W_2*: the second ball drawn is white,*
B_3*: the third ball selected is black,* B_4*: the fourth ball selected is black.*
We are interested in the probability of $\mathbb{P}(W_1, W_2, B_3, B_4)$.

Answer 1. *Note that*

$$\mathbb{P}(W_1, W_2, B_3, B_4) = \mathbb{P}(B_4, B_3, W_2, W_1)$$

Therefore, we can write

$$\mathbb{P}(B_4, B_3, W_2, W_1) = \mathbb{P}(B_4|B_3, W_2, W_1)\mathbb{P}(B_3, W_2, W_1)$$

and continue decomposing the last part as

$$\mathbb{P}(B_3, W_2, W_1) = \mathbb{P}(B_3|W_2, W_1)\mathbb{P}(W_2, W_1)$$

and again

$$\mathbb{P}(W_2, W_1) = \mathbb{P}(W_1)\mathbb{P}(W_2|W_1).$$

Putting them altogether:

$$\mathbb{P}(B_4, B_3, W_2, W_1) = \mathbb{P}(B_4|B_3, W_2, W_1)\mathbb{P}(B_3|W_2, W_1)\mathbb{P}(W_2|W_1)\mathbb{P}(W_1).$$

Since we are sampling without replacement, it is easy to observe that

$$\mathbb{P}(B_4, B_3, W_2, W_1) = \frac{8}{12} \cdot \frac{9}{13} \cdot \frac{5}{14} \cdot \frac{6}{15}$$

Answer 2.

$$\frac{\binom{6}{1}\binom{9}{0}}{\binom{15}{1}} \cdot \frac{\binom{5}{1}\binom{9}{0}}{\binom{14}{1}} \cdot \frac{\binom{9}{1}\binom{4}{0}}{\binom{13}{1}} \cdot \frac{\binom{8}{1}\binom{4}{0}}{\binom{12}{1}} = \frac{6}{15} \cdot \frac{5}{14} \cdot \frac{9}{13} \cdot \frac{8}{12}$$

□

Example 2.14. *Suppose an urn contains 8 red balls and 4 white balls. We draw 2 balls from the urn without replacement. If we assume that as we draw the each ball in the urn is equally likely to be drawn, what is the probability that both balls drawn are red?*

Let R_1 *and* R_2 *denote respectively the events that the first and the second ball drawn is red. Now given that the first ball selected is red, there 7 remaining red balls and 4 white balls and so* $\mathbb{P}(\mathrm{R}_2|\mathrm{R}_1) = 7/11$*. As* $\mathbb{P}(\mathrm{R}_1) = 8/12$*, the desired probability is*

$$\mathbb{P}(\mathrm{R}_1 \cap \mathrm{R}_2) = \mathbb{P}(\mathrm{R}_1)\mathbb{P}(\mathrm{R}_2|\mathrm{R}_1) = (8/12)(7/11) = 14/33.$$

Note that another way to work this problem is use combinations by thinking of drawing a pair of balls in quick succession:

$$\mathbb{P}(\mathrm{R}_1 \cap \mathrm{R}_2) = \frac{\binom{8}{1}\binom{4}{0}}{\binom{12}{1}} \cdot \frac{\binom{7}{1}\binom{4}{0}}{\binom{11}{1}} = \frac{8}{12} \cdot \frac{7}{11} = \frac{14}{33}.$$

□

Chapter 3

Conditional Probability and Independence

What is the probability that the outcome of a dice roll is 2 given that the outcome is an even number? Many times we need to compute probabilities under the assumption that some event has occurred. We deal with this type of experiments using conditional probabilities. Independence of events is also an important concept in probability. When two events are independent, the outcome of one experiment does not affect the likelihood of occurring for the other event. Observing 10 heads in 10 trials of a fair coin flip does not make the next trial to be more likely a tail. Because, coin flips are independent trials. This chapter introduces the concept of conditional probability and elaborates conditions that are required for independence of events.

3.1 Conditional Probability

The conditional probability $\mathbb{P}(A|B)$ is the probability that A occurs given that B occurs. The vertical bar | is read as "conditioned on" or "given". In conditional probability $\mathbb{P}(A|B)$, the probability space is restricted to B and we are interested in the relative probability assigned to the portion of A that is contained in B.

Definition 12 (Conditional probability)**.** *Given two events* A and B with $\mathbb{P}(B) > 0$, the conditional probability of A given B has occurred is defined as

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

Proposition 10 (Multiplication rule)**.**

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \times \mathbb{P}(B).$$

Proposition 11 (Law of total probability)**.** *Let* $A_1, ..., A_n$ *be mutually exclusive and exhaustive events. Then for any other event* B

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

Proof. Since the event B satisfies the equality

$$B = (B \cap A_1) \cup (B \cap A_2) \cup ... \cup (B \cap A_n).$$

The probability $\mathbb{P}(B)$ can be written as

$$\mathbb{P}(B) = \mathbb{P}\left\{(B \cap A_1) \cup (B \cap A_2) \cup ... \cup (B \cap A_n)\right\}.$$

Since A_i's are exclusive and exhaustive events, the sets $B \cap A_i$ are mutually exclusive.

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

Where in the last step, we applied the multiplication rule. □

Remark 3. *Notice that a simplified version of law of total probability can be written as:*

$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c).$$

Based on the law of total probability, the probability of an event B is a weighted average of the conditional probability of B given that event A_i has occurred over all the possibilities of A_i.

3.2 Bayes' theorem

Bayes' Theorem[1] for events gives the conditional probability $\mathbb{P}(A|B)$ in terms of the conditional probability $\mathbb{P}(B|A)$.

[1] Thomas Bayes (1702-1761)

Theorem 2 (Bayes' theorem)**.** *Let* $\mathrm{A}_1, ..., \mathrm{A}_\mathrm{n}$ *be mutually exclusive and exhaustive events with* $\mathbb{P}(\mathrm{A}_\mathrm{i}) > 0$ *for all* $i = 1, ..., n$*. Then for any other event* B with $\mathbb{P}(\mathrm{B}) > 0$

$$\mathbb{P}(\mathrm{A_j}|\mathrm{B}) = \frac{\mathbb{P}(\mathrm{A_j} \cap \mathrm{B})}{\mathbb{P}(\mathrm{B})} = \frac{\mathbb{P}(\mathrm{B}|\mathrm{A_j})\mathbb{P}(\mathrm{A_j})}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathbb{P}(\mathrm{B}|\mathrm{A_i})\mathbb{P}(\mathrm{A_i})}, \quad \mathrm{j} = 1, ..., \mathrm{n}.$$

Similar to the law of total probability, Bayes' theorem in the simple form can be written as

$$\mathbb{P}(\mathrm{A}|\mathrm{B}) = \frac{\mathbb{P}(\mathrm{B}|\mathrm{A})\mathbb{P}(\mathrm{A})}{\mathbb{P}(\mathrm{B}|\mathrm{A})\mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{B}|\mathrm{A^c})\mathbb{P}(\mathrm{A^c})}.$$

for $\mathbb{P}(\mathrm{B}) > 0$.

Proof. Notice that

$$\mathbb{P}(\mathrm{A} \cap \mathrm{B}) = \mathbb{P}(\mathrm{B} \cap \mathrm{A})$$

Using the multiplication rule

$$\mathbb{P}(\mathrm{A}|\mathrm{B})\mathbb{P}(\mathrm{B}) = \mathbb{P}(\mathrm{B}|\mathrm{A})\mathbb{P}(\mathrm{A})$$

Therefore,

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

Applying the law of total probability in the denominator proves the result. □

Applications of Bayes' theorem are versatile. A famous example of such applications is related to the interpretation of screening tests used to detect certain types of diseases such as cancer. Often screening tests are characterized by their *sensitivity* and *specificity* scores, which are widely used metrics for measuring the performance of a test. Sensitivity of a test refers to the probability that the test correctly identifies a patient as *diseased* when the patient has in fact the disease. Specificity refers to the probability that the test correctly identifies a patient as *healthy* when the patient is in fact disease-free. In many rare diseases, the physician may be interested in knowing the probability that the patient has the disease given that the result of the screening test is positive.

Example 3.1. *A population of 10000 patient is selected to study the performance of a mammography screening test. Using a perfectly accurate Biopsy test, it was revealed that only 10 of them had early stage breast cancer. Those patients then underwent a mammography test. The results of the screening test are presented in Table 3.1.*

	Diseased	Healthy
Test is positive	9	50
Test is negative	1	9940

Table 3.1: Result of mammography screening test for 10000 patients.

During the annual population level screening for breast cancer, the mammography test of a patient turned out to be positive. The physician who had access to the data on the performance of mammography test decided to compute Prob{Patient has breast cancer|Mammography test is positive}.

To answer this question we use the following notation:

Events: $T+$: Test positive, $T-$: Test negative, D: Diseased, H: Healthy. From the table we can compute the joint distributions:

$\mathbb{P}(T+\cap D) = 0.0009$, $\mathbb{P}(T+\cap H) = 0.005$, $\mathbb{P}(T-\cap D) = 0.0001$, *and* $\mathbb{P}(T-\cap H) = 0.994$.

To compute the marginal probabilities, we apply the law of total probability.

$$\begin{aligned}
\mathbb{P}(T+) &= \mathbb{P}(T+\cap D) + \mathbb{P}(T+\cap H) = 0.0059. \\
\mathbb{P}(T-) &= \mathbb{P}(T-\cap D) + \mathbb{P}(T-\cap H) = 0.9941 \\
\mathbb{P}(D) &= \mathbb{P}(D\cap T+) + \mathbb{P}(D\cap T-) = 0.001 \\
\mathbb{P}(H) &= \mathbb{P}(H\cap T+) + \mathbb{P}(H\cap T-) = 0.999
\end{aligned}$$

It is also easy to observe that $\mathbb{P}(T+|D) = 0.9$. Notice that since it is a conditional probability, the universe is restricted to the cases which are Diseased.

Applying Bayes' theorem, we obtain

$$\mathbb{P}(D|T+) \quad = \quad \frac{\mathbb{P}(D\cap T+)}{\mathbb{P}(T+)} = \frac{.0009}{.0059} = 0.15\%$$

□

The low probability of being cancerous given mammography was positive infers that additional evidence should be taken into account before making concrete conclusions about the presence of disease even if the test sensitivity to find the positive cases is high.

Example 3.2 (Monty Hall Problem)**.** *The classical version of the problem is described as follows: A contestant is shown three identical doors. Behind one of them is a car. The other two conceal goats. The contestant is asked to choose, but not open, one of the doors. After doing so, Monty, who knows where the car hides, opens one of the two remaining doors. He always opens a door he knows to be incorrect (goat-concealing doors will be referred to as the incorrect doors), and randomly chooses which door to open when he has more than one option (which happens on the occasion where the contestant's initial choice conceals the car). After opening an incorrect door, Monty gives the contestant the option of either switching to the other unopened door or sticking with their original choice. The contestant then receives whatever is behind the door they choose. What should the contestant do?*

Initially when the contestant chooses the door, he has a $\frac{1}{3}$ chance of picking the car. This must mean that the other doors combined have a $\frac{2}{3}$ chance of winning. But after Carol opens a door with a goat behind it, how do the probabilities change? Well, everyone knows that there is a goat behind one of the doors that the contestant did not pick. So no matter whether the contestant is winning or not, Carol is always able to open one of the other doors to reveal a goat. This means that the contestant still has a $\frac{1}{3}$ chance of winning. Also the door that Carol opened has no chance of winning. What about the last door? It must have a $\frac{2}{3}$ chance of containing the car, and so the contestant has a higher chance of winning if he or she switches doors.

We will denote C_i the event that the car is behind door i and M_i the event that Monty opens door i. Now imagine a modified version of this problem with the following information:

$$\mathbb{P}(\mathrm{M}_2|\mathrm{C}_1) = 1/2,$$

$$\mathbb{P}(\mathrm{M}_2|\mathrm{C}_2) = 0,$$

$$\mathbb{P}(\mathrm{M}_2|\mathrm{C}_3) = 1,$$

The prior probability that the car is behind any door is $1/3$*. That is* $\mathbb{P}(\mathrm{C}_1) = \mathbb{P}(\mathrm{C}_2) = \mathbb{P}(\mathrm{C}_3) = 1/3$.
Suppose the contestant picks door 1 *and Monty opens door* 2*. We are interested in the probability that the car is behind the any of the doors (*1 *and* 3*) given that Manty opened* 2.

By law of total probability, the probability that Monty opens door 2 *is*

$$\begin{aligned}\mathbb{P}(\mathrm{M}_2) &= \mathbb{P}(\mathrm{M}_2|\mathrm{A}) \times \mathbb{P}(\mathrm{A}) + \mathbb{P}(\mathrm{M}_2|\mathrm{B}) \times \mathbb{P}(\mathrm{B}) + \mathbb{P}(\mathrm{M}_2|\mathrm{C}) \times \mathbb{P}(\mathrm{C}) \\ &= 1/6 + 0 + 1/3 = 1/2.\end{aligned}$$

By Bayes rule

$$\begin{aligned}\mathbb{P}(\mathrm{C}_1|\mathrm{M}_2) &= \frac{\mathbb{P}(\mathrm{M}_2|\mathrm{C}_1) \times \mathbb{P}(\mathrm{C}_1)}{\mathbb{P}(\mathrm{M}_2)} \\ &= (1/6)/(1/2) = 1/3 \\ \mathbb{P}(\mathrm{C}_3|\mathrm{M}_2) &= \frac{\mathbb{P}(\mathrm{M}_2|\mathrm{C}_3) \times \mathbb{P}(\mathrm{C}_3)}{\mathbb{P}(\mathrm{M}_2)} \\ &= (1/3)/(1/2) = 2/3.\end{aligned}$$

This example shows that given the circumstances, it is wise for the contestant to change his choice.

□

Example 3.3. *In a city 95% of the cabs are green and 5% are blue. A colorblind eyewitness observes a hit-and-run cab accident, and reported the car involved in the accident to be blue. Further investigations showed that the eyewitness correctly identifies the colors only 80% of the time. What is the probability that the cab actually was blue?*
Let B *and* G *represent the event that the color of the car involved was Blue and Green cab respectively. Let* b *and* g *be the event that the eyewitness reported a Blue and Green cab respectively. Using these notations,* $\mathbb{P}(b|B) = \mathbb{P}(g|G) = 0.8$ *and therefore* $\mathbb{P}(b|G) = 0.2$*. Using Bayes' theorem,*

$$\begin{aligned}\mathbb{P}(B|b) &= \frac{\mathbb{P}(b|B) \cdot \mathbb{P}(B)}{\mathbb{P}(b|B) \cdot \mathbb{P}(B) + \mathbb{P}(b|G) \cdot \mathbb{P}(G)} \\ &= \frac{0.8 \cdot 0.05}{0.8 \cdot 0.05 + 0.2 \cdot 0.95} = 0.173.\end{aligned}$$

□

Example 3.4. [2] *Stores A, B and C have* 50, 75 *and* 100 *employees respectively. and respectively, 50%, 60% and 70% of the employees are women. Resignations are equally likely among all employees, regardless of sex. One employee resigns and this is a woman. What is the probability that she works in store A?*

Let W be the event that a woman employee resigns from anywhere, and let A, B and C denote the event that a randomly selected employee works at the respective store. Then $\mathbb{P}(A) = 50/225$, $\mathbb{P}(B) = 75/225$ *and* $\mathbb{P}(C) = 100/225$. *Likewise the probabilities of resignation of a woman from a store is given by the information to be* $\mathbb{P}(W|A) = 0.50$, $\mathbb{P}(W|B) = 0.60$, *and* $\mathbb{P}(W|C) = 0.70$. *Then we can use Bayes Theorem (re-deriving it in the process of using it):*

$$
\begin{aligned}
\mathbb{P}(A|W) &= \frac{\mathbb{P}(A \cap W)}{\mathbb{P}(W)} \\
&= \frac{\mathbb{P}(W|C)\mathbb{P}(C)}{\mathbb{P}(W|A)\mathbb{P}(A) + \mathbb{P}(W|B)\mathbb{P}(B) + \mathbb{P}(W|C)\mathbb{P}(C)} \\
&= \frac{(0.50)(50/225)}{(0.50)(50/225) + (0.60)(75/225) + (0.70)(100/225)} \\
&= 0.17857
\end{aligned}
$$

Example 3.5. *In answering a question on a multiple choice test, a student either knows the answer or the student just guesses. Suppose that the probability that the student knows the answer is* 0.75, *and the probability the student doesn't know the answer and guesses is* 0.25. *Assuming that there are* 5 *choices for each multiple-choice question, then we take the probability that the student who guesses will be correct is* $1/5 = 0.20$. *What is the conditional probability that the student knew the answer to a question given that the student answered it correctly?*

Let C be the even that the student answers the question correctly. Let K be the probability the student knows the answer. $\mathbb{P}(C|K) = 1$.

---

[2] Adapted from *A First Course in Probability* by Sheldon Ross, Macmillan, 1976, Chapter 3, Problem 18, page 80.

$$\begin{aligned}
\mathbb{P}(K|C) &= \frac{\mathbb{P}(K\cap C)}{\mathbb{P}(C)} \\
&= \frac{\mathbb{P}(C|K)\mathbb{P}(K)}{\mathbb{P}(C|K)\mathbb{P}(K)+\mathbb{P}(C|K^c)\mathbb{P}(K^c)} \\
&= \frac{0.75}{0.75\cdot 1+0.25\cdot 0.20} \\
&= 0.9375
\end{aligned}$$

3.2.1 A side Note About Bayesian Inference

Bayesian inference is one of the most popular techniques in Bayesian statistics to make inference about the distributions. To clarify, suppose we guess that the distribution of a population is normal. This wild guess is our *prior* knowledge about the underlying distribution of the population. We start too draw samples from the population and based on the drawn samples we update our assumption about the population distribution. Our updated belief is called *posterior.*

In the Bayes' theorem, if we let B to be some data and A to be a model, we can use the probability Prob(data|model) from the statistical model to obtain the inference Prob(model|data).

$$\text{Prob(model|data)} = \frac{\text{Prob(data|model) Prob(model)}}{\text{Prob(data)}}.$$

Where, Prob(data|model) is called *likelihood*, Prob(model) is called *prior* and Prob(model|data) is called *posterior.*

Now that we've introduced the notion of conditional probability, we can see how it is used in real world settings. Conditional probability is at the heart of a subject called *Bayesian inference*, used extensively in fields such as machine learning, communications and signal processing. Bayesian inference is a way to *update knowledge* after making an observation. For example, we may have an estimate of the probability of a given event A. After event B occurs, we can update this estimate to $\mathbb{P}(\text{A}|\text{B})$. In this interpretation, $\mathbb{P}(A)$ can be thought of as a *prior* probability: our assessment of the likelihood of an event of interest A *before* making an observation. It reflects our prior knowledge. $\mathbb{P}(\text{A}|\text{B})$ can be interpreted as the *posterior* probability of A after the observation. It reflects our new knowledge.

3.3 Independence Of Events

In order to clarify the concept of independence, think of the following example. In a coin flip experiment, you flip a fair coin 10 times and you observe all of them as tails. You start to wonder whether observing 10 tails makes the next flip more likely to be a head. In fact the answer is no. No matter how many tails you observed, each flip has exactly 50% chance of being a head. This implies that in coin flip experiment the outcomes are independent. One other way to visualize this experiment is as follows. Suppose a box contains two one sided coins. Each coin has only one side, one coin is head and the other is a tail. Each time you draw one coin randomly and after recording the coin type, you return the coin into the box. This experiment is in fact equivalent to the fair coin flip where each side is equally likely. Since in each draw, the sample space remains the same, i.e., $\Omega = \{\mathrm{H}, \mathrm{T}\}$, it is called experiment with replacement. A dice roll is also another example of experiment in which the outcomes in each roll are independent.

On the other hand, in the experiment of dealing cards, if all of the first 10 cards turn out to be diamonds, what can we say about the probability that the next card will be a diamond? Obviously, in a deck of 52 cards, given that the first 10 were all diamonds makes the next card less likely to be a diamond. Before the experiment begins the probability is $13/52 = 0.25$, whereas after first 10 appearance of diamonds, only 3 diamonds remain among the remaining 42 cards, making the probability of the next one to be a diamond as small as $3/42 = 0.07$. In other words, in this experiment the outcomes in each dealing of the card is no longer independent of the previous cards. It is easy to observe that in this experiment the sample space does not remain the same after each dealing of the cards. This is an example of experiments without replacement.

Definition 13 (Independent Events)**.** *Two events* A and B are independent if $\mathbb{P}(\mathrm{A}|\mathrm{B}) = \mathbb{P}(\mathrm{A})$.

Note that if B is independent of A, then A is independent of B.

Proposition 12. A and B are independent if and only if

$$\mathbb{P}(\mathrm{A} \cap \mathrm{B}) = \mathbb{P}(\mathrm{A}) \times \mathbb{P}(\mathrm{B}).$$

Proof. From the multiplicative rule,

$$\mathbb{P}(\mathrm{A} \cap \mathrm{B}) = \mathbb{P}(\mathrm{A}|\mathrm{B}) \times \mathbb{P}(\mathrm{B})$$

Since A and B are independent, by definition $\mathbb{P}(\mathrm{A}|\mathrm{B}) = \mathbb{P}(\mathrm{A})$, which proves the result.

$$\begin{aligned} \mathbb{P}(\mathrm{A}) \times \mathbb{P}(\mathrm{B}) &= \mathbb{P}(\mathrm{A}|\mathrm{B}) \times \mathbb{P}(\mathrm{B}) \\ &= \mathbb{P}(\mathrm{A} \cap \mathrm{B}) \end{aligned}$$

Where the first equality is by definition of independence. □

Definition 14. *Events* $\mathrm{A}_1, ..., \mathrm{A}_\mathrm{n}$ *are mutually independent if for every* $k = 2, 3, \ldots, n$ *and every subset of indices* $i_1, i_2, ..., i_k$

$$\mathbb{P}(\mathrm{A}_{\mathrm{i}_1} \cap \mathrm{A}_{\mathrm{i}_2} \cap ...\mathrm{A}_{\mathrm{i}_k}) = \mathbb{P}(\mathrm{A}_{\mathrm{i}_1}) \times \mathbb{P}(\mathrm{A}_{\mathrm{i}_2}) \times ...\mathbb{P}(\mathrm{A}_{\mathrm{i}_k}).$$

As an example, suppose $n = 3$, which produces three events $\mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3$. Also $n = 3$ results in $k = 2, 3$.
When $k = 2$,

$$\begin{aligned} \mathbb{P}(\mathrm{A}_1 \cap \mathrm{A}_2) = \mathbb{P}(\mathrm{A}_1) \times \mathbb{P}(\mathrm{A}_2) \\ \mathbb{P}(\mathrm{A}_1 \cap \mathrm{A}_3) = \mathbb{P}(\mathrm{A}_1) \times \mathbb{P}(\mathrm{A}_3) \\ \mathbb{P}(\mathrm{A}_2 \cap \mathrm{A}_3) = \mathbb{P}(\mathrm{A}_2) \times \mathbb{P}(\mathrm{A}_3) \end{aligned}$$

When $k = 3$,

$$\begin{aligned} \mathbb{P}(\mathrm{A}_1 \cap \mathrm{A}_2 \cap \mathrm{A}_3) &= \mathbb{P}(\mathrm{A}_1) \times \mathbb{P}(\mathrm{A}_2) \times \mathbb{P}(\mathrm{A}_3) \\ &= \mathbb{P}(\mathrm{A}_1 \cap \mathrm{A}_2) \times \mathbb{P}(\mathrm{A}_3). \\ &= \mathbb{P}(\mathrm{A}_1 \cap \mathrm{A}_3) \times \mathbb{P}(\mathrm{A}_2). \\ &= \mathbb{P}(\mathrm{A}_2 \cap \mathrm{A}_3) \times \mathbb{P}(\mathrm{A}_1). \end{aligned}$$

Definition 15 (Conditional independence)**.** *Events* A and B are said to be conditionally independent given C when

$$\mathbb{P}(\mathrm{A} \cap \mathrm{B}|\mathrm{C}) = \mathbb{P}(\mathrm{A}|\mathrm{C})\mathbb{P}(\mathrm{B}|\mathrm{C}).$$

Notice that

$$\begin{aligned} \mathbb{P}(\mathrm{A} \cap \mathrm{B}|\mathrm{C}) &= \frac{\mathbb{P}(\mathrm{A} \cap \mathrm{B} \cap \mathrm{C})}{\mathbb{P}(\mathrm{C})} \\ &= \frac{\mathbb{P}(\mathrm{A}|\mathrm{B} \cap \mathrm{C})\mathbb{P}(\mathrm{B} \cap \mathrm{C})}{\mathbb{P}(\mathrm{C})} \\ &= \frac{\mathbb{P}(\mathrm{A}|\mathrm{B} \cap \mathrm{C})\mathbb{P}(\mathrm{B}|\mathrm{C})\mathbb{P}(\mathrm{C})}{\mathbb{P}(\mathrm{C})} \\ &= \mathbb{P}(\mathrm{A}|\mathrm{B} \cap \mathrm{C})\mathbb{P}(\mathrm{B}|\mathrm{C}) \\ &= \mathbb{P}(\mathrm{A}|\mathrm{C})\mathbb{P}(\mathrm{B}|\mathrm{C}) \end{aligned}$$

where the last equality follows from the fact that given C, A is conditionally independent of B, i.e., $\mathbb{P}(A|B \cap C) = \mathbb{P}(A|C)$.