Measure Theoretic Probability

Prerequisites: Real Analysis, Classical Probability

Sagar Udasi

November 2, 2024

Contents

Chapter 1

Introduction to σ-algebra

In the classical probability, we encountered *Bertnard's Paradox*, which highlighted the significance of rotational and translational invariance in probability measurements. We discovered that probabilities, much like lengths, areas, or volumes, should remain unchanged when subjected to such transformations. For example, if two points are separated by a distance d, shifting them by an equal amount in the same direction preserves that distance. This invariance hints that probability is not merely a tool for quantifying uncertainty but rather a form of measure - just like length, area or volume.

This realization serves as our starting point for a deeper, more formal approach to understanding probability, known as **Measure Theoretic Probability**. By treating probability as a measure, we establish a rigorous mathematical foundation that allows us to precisely define, manipulate, and compute probabilities, even in complex scenarios involving infinite spaces or continuous distributions.

In this chapter, we will build from the fundamentals of measure theory, gradually developing the key concepts required to form a robust understanding of probability in this framework [1].

1.1 Introduction to $\sigma - algebra$

In the early chapters of *Real Analyis*, we introduced the concept of a *field*. A field is an ordered triple, for example, $(\mathbb{Q}, +, \times)$, consisting of the set of rational numbers $\mathbb{Q}$ and two binary operations, $+$ and $\times$, defined on them. These operations follow specific properties, such as having an identity element, the existence of an inverse element for each non-zero element, and distributivity, among others. This structure forms what is commonly referred to as *arithmetic or numeric algebra*.

But what if we change the set and the operations? Suppose instead of $\mathbb{Q}$, we take the set $\{0, 1\}^\infty$ (the set of all binary sequences) and define appropriate binary operations, such as $+$ and $\cdot$. The resulting structure is called a *boolean algebra*. Similarly, if we take the set of matrices M and define addition and multiplication operations on them, we obtain what is known as *matrix algebra*. These examples illustrate that the notion of algebra is not restricted to numbers; it can be generalized to other sets with appropriate operations.

Now, consider a large set with many subsets as its elements, and define two operations: union $\cup$ and intersection $\cap$. This leads to what is known as a *set algebra*, which is central to our discussion.

Definition 1.1. *Let Ω be a sample space and let $\mathcal{F}_0$ be a collection of subsets of Ω. Then, $\mathcal{F}_0$ is said to be an algebra (or a field) if the following conditions hold:*

1. $\emptyset \in \mathcal{F}_0$.
2. *If $A \in \mathcal{F}_0$, then $A^c \in \mathcal{F}_0$.*
3. *If $A \in \mathcal{F}_0$ and $B \in \mathcal{F}_0$, then $A \cup B \in \mathcal{F}_0$.*

While the terms *field* and *algebra* are sometimes used interchangeably in the context of sets, there is a subtle difference when we generalize to other structures. A *field* refers specifically to a set with two binary operations (like $+$ and $\times$) that satisfy a complete set of properties such as associativity, commutativity, distributivity, and the existence of identity and inverse elements.

On the other hand, an *algebra* is a broader concept. It is a structure consisting of a set and operations that may or may not satisfy all the properties required of a field. For instance, in set theory, a set algebra satisfies closure under union, intersection, and complement, but it does not necessarily satisfy all the numeric properties of a field, such as the existence of multiplicative inverses. Thus, while all fields can be considered a type of algebra, not all algebras are fields. The key distinction lies in the specific operations and properties defined on the set.

Theorem 1.1. *An algebra is closed under finite union and finite intersection.*

Proof. **Closed Under Finite Union:**

To prove that $\mathcal{F}_0$ is closed under finite union, we proceed by induction.

Base Case: Let $A_1, A_2 \in \mathcal{F}_0$. By definition of an algebra, $A_1 \cup A_2 \in \mathcal{F}_0$. This shows that the union of two sets in $\mathcal{F}_0$ is also in $\mathcal{F}_0$.

Induction Step: Suppose for some $n \in \mathbb{N}$, the union of n sets in $\mathcal{F}_0$, say $A_1, A_2, \ldots, A_n$, is also in $\mathcal{F}_0$. That is,

$$A_1 \cup A_2 \cup \ldots \cup A_n \in \mathcal{F}_0.$$

Now, consider $A_1 \cup A_2 \cup \ldots \cup A_n \cup A_{n+1}$. We can rewrite this as:

$$(A_1 \cup A_2 \cup \ldots \cup A_n) \cup A_{n+1}.$$

By the induction hypothesis, $(A_1 \cup A_2 \cup \ldots \cup A_n) \in \mathcal{F}_0$. Since $A_{n+1} \in \mathcal{F}_0$ and $\mathcal{F}_0$ is closed under the union of two sets, it follows that:

$$(A_1 \cup A_2 \cup \ldots \cup A_n) \cup A_{n+1} \in \mathcal{F}_0.$$

By the principle of mathematical induction, $\mathcal{F}_0$ is closed under finite union.

Closed Under Finite Intersection:

To show closure under finite intersection, note that for any sets $A, B \in \mathcal{F}_0$, we have $A^c, B^c \in \mathcal{F}_0$ because complements of sets in an algebra are also in the algebra.

Using De Morgan's laws, we know that:

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

Since $A^c, B^c \in \mathcal{F}_0$ and $\mathcal{F}_0$ is closed under finite union, it follows that $A^c \cup B^c \in \mathcal{F}_0$. Therefore, $(A^c \cup B^c)^c \in \mathcal{F}_0$, meaning $A \cap B \in \mathcal{F}_0$.

By similar reasoning and using induction, it can be shown that $\mathcal{F}_0$ is closed under the intersection of any finite number of sets. Thus, $\mathcal{F}_0$ is closed under finite intersection.

□

We have not defined the concept of *event* yet. Informally, for now consider that an event is an subset of sample space that is of our interest. A natural question that arises at this point is *Is the structure of an algebra enough to study events of typical interest in probability theory?*

An Event Not Included in an Algebra

Consider the following example. Toss a coin repeatedly until the first heads appears. The sample space is:

$$\Omega = \{H, TH, TTH, \dots\}$$

where H represents heads appearing on the first toss, TH represents tails followed by heads, TTH represents two tails followed by heads, and so on.

Now, suppose we are interested in determining whether the number of tosses before seeing a head is even. Let E denote this event. Then,

$$E = \{TH, TTTH, TTTTTH, \dots\}$$

which includes all outcomes where heads appears after an even number of tosses.

Notice that E is a countably infinite union of individual outcomes:

$$E = \{TH\} \cup \{TTTH\} \cup \{TTTTTH\} \cup \dots$$

However, an *algebra* is defined to contain only finite unions of subsets. Since E involves a countably infinite union, it cannot be part of the algebra of subsets of Ω. This shows that our *event* of interest is not included in the algebra.

This limitation motivates the need for a more comprehensive structure called a *σ-algebra*. A σ-algebra extends the notion of an algebra by allowing countably infinite unions of subsets, ensuring that events like E are included within the framework of probability theory.

Definition 1.2. *A collection $\mathcal{F}$ of subsets of Ω is called a σ-algebra (or σ-field) if:*

1. $\emptyset \in \mathcal{F}$.
2. *If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ (i.e., the complement of A is also in $\mathcal{F}$).*
3. *If $A_1, A_2, A_3, \ldots$ is a countable collection of subsets in $\mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

Note that, unlike an algebra, a σ-algebra is closed under countable union and countable intersection.

Examples of $\sigma - algebras$**:**

Here are some intuitive examples of σ-algebras:

1. **Trivial σ-algebra:** The smallest σ-algebra on a sample space Ω is $\mathcal{F} = \{\emptyset, \Omega\}$. This is known as the trivial σ-algebra and contains only the empty set and the entire sample space.
2. **Power Set σ-algebra:** The largest σ-algebra on a sample space Ω is the power set of Ω, denoted as 2^{Ω}. It includes all possible subsets of Ω. This is the most comprehensive σ-algebra possible on Ω.
3. **Finite and Countable σ-algebras:** Consider a finite or countable sample space, such as $\Omega = \{1, 2, 3, \ldots\}$. The collection of all subsets of Ω forms a σ-algebra, as it is closed under countable unions, intersections, and complements.

Theorem 1.2. *Every σ-algebra is an algebra, but the converse is not true.*

Proof. **Part 1: Every σ-algebra is an algebra**

Let $\mathcal{F}$ be a σ-algebra.

1. **Contains the empty set:** By the definition of a σ-algebra, we have $\emptyset \in \mathcal{F}$.
2. **Closed under complementation:** If $A \in \mathcal{F}$, then by definition, $A^c \in \mathcal{F}$.
3. **Closed under finite unions:** Let $A, B \in \mathcal{F}$. We can consider the finite union:

$$A \cup B = A \cup B = \bigcup_{i=1}^{2} A_i.$$

   Here, we can denote $A_1 = A$ and $A_2 = B$. Since $\mathcal{F}$ is closed under countable unions, we have:

$$A \cup B \in \mathcal{F}.$$

Since $\mathcal{F}$ satisfies all three properties of an algebra, we conclude that every σ-algebra is indeed an algebra.

Part 2: The converse is not true

To show that not every algebra is a σ-algebra, we can provide a counterexample.

Consider the set $\Omega = \{1,2,3\}$ and the algebra $\mathcal{A} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$.

1. **Contains the empty set:** $\emptyset \in \mathcal{A}$.
2. **Closed under complementation:** The complement of each set in $\mathcal{A}$ is also in $\mathcal{A}$.
3. **Closed under finite unions:** The union of any finite number of sets in $\mathcal{A}$ is also in $\mathcal{A}$.

However, the collection $\mathcal{A}$ is not a σ-algebra because it is not closed under countable unions. For instance, if we consider the countable collection of subsets:

$$A_1 = \{1\}, \quad A_2 = \{2\}, \quad A_3 = \{3\}, \quad \dots$$

the union $\bigcup_{i=1}^{\infty} A_i = \{1,2,3\} = \Omega$, which is included, but if we consider an infinite union of disjoint sets from $\mathcal{A}$ that leads to more than three elements, it will not be contained within $\mathcal{A}$.

Thus, we conclude that not every algebra is a σ-algebra. $\square$

Examples of algebras which are not σ-algebras:

Below are some simple examples of an algebra that is not a σ-algebra:

Example 1: The Finite Subsets of $\mathbb{N}$

Consider the set $\Omega = \mathbb{N}$, the set of all natural numbers. Let $\mathcal{A}$ be the collection of all finite subsets of $\mathbb{N}$ along with $\mathbb{N}$ itself. This collection forms an **algebra** because:

- The union or intersection of any two finite sets is finite (or possibly $\mathbb{N}$).
- The complement of any finite subset is also an infinite subset, and in this case, it is $\mathbb{N}$ (which belongs to $\mathcal{A}$).

However, $\mathcal{A}$ is **not a σ-algebra** because it is not closed under countable union. For instance, if we take a sequence of singletons $\{1\}, \{2\}, \{3\}, \dots$, the union of these singletons is $\mathbb{N}$, which is an infinite set. While $\mathbb{N}$ is in $\mathcal{A}$, the complement of this countable union would not necessarily belong to $\mathcal{A}$, as it may not be finite.

Example 2: Intervals on the Real Line

Consider $\Omega = [0,1]$ and let $\mathcal{A}$ be the collection of all finite unions of intervals of the form $[a,b]$, where $0 \le a \le b \le 1$. This collection $\mathcal{A}$ forms an **algebra** because:

- The union and intersection of a finite number of intervals of this form are again finite unions of intervals of this form.
- The complement of a finite union of such intervals is also a finite union of intervals.

However, $\mathcal{A}$ is **not a σ-algebra** because it is not necessarily closed under countable unions. For example, if we take a sequence of intervals $\left[0, \frac{1}{2}\right], \left[\frac{1}{2}, \frac{3}{4}\right], \left[\frac{3}{4}, \frac{7}{8}\right], \ldots$ such that they cover $[0, 1]$ as a whole, their countable union would be $[0, 1]$. Although $[0, 1]$ is in $\mathcal{A}$, the structure of $\mathcal{A}$ doesn't guarantee closure under all such countable unions.

Example 3: The Power Set of a Finite Set

Let $\Omega = \{a, b, c\}$ be a finite set. The collection $\mathcal{A}$ of all subsets of Ω (also known as the power set of Ω) forms an **algebra** because:

- Any union, intersection, or complement of subsets of a finite set remains a subset of that finite set.

However, even though this is a trivial example, it demonstrates that an algebra is not necessarily a σ-algebra because σ-algebras are designed to handle infinite cases. In this finite scenario, $\mathcal{A}$ satisfies the properties of both an algebra and a σ-algebra, but it shows that if the set Ω were infinite, $\mathcal{A}$ would not generally be closed under countable operations.

Exercise 1.1. *Consider the random experiment of throwing a die. If a statistician is interested in the occurrence of either an odd or an even outcome, construct a sample space and a σ-algebra of subsets of this sample space.*

Solution 1.1. ***Sample Space (Ω)***: *The sample space consists of all possible outcomes when throwing a six-sided die. Therefore, we can define the sample space as:*

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Events of Interest: *The statistician is interested in the occurrence of either an odd or an even outcome. We can categorize the outcomes as follows:*

- ***Odd Outcomes***: $\{1, 3, 5\}$
- ***Even Outcomes***: $\{2, 4, 6\}$

Constructing the σ-Algebra ($\mathcal{F}$): *A σ-algebra is a collection of subsets of Ω that satisfies the following properties:*

- *It contains the empty set and the sample space itself.*
- *It is closed under complementation.*
- *It is closed under countable unions.*

Given the events of interest, we can construct the σ-algebra as follows:

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 3, 5\}, \{2, 4, 6\}, \{1, 2, 3, 4, 5, 6\}\}$$

Checking the Properties of the σ-Algebra:

Contains the Empty Set and Sample Space: *$\emptyset \in \mathcal{F}$ and $\Omega = \{1,2,3,4,5,6\} \in \mathcal{F}$.*

Closed under Complementation:

- *The complement of $\emptyset$ is $\{1,2,3,4,5,6\}$, which is in $\mathcal{F}$.*
- *The complement of $\{1,3,5\}$ is $\{2,4,6\}$, which is in $\mathcal{F}$.*
- *The complement of $\{2,4,6\}$ is $\{1,3,5\}$, which is in $\mathcal{F}$.*

Closed under Countable Unions: *For any events in $\mathcal{F}$, the union will also be in $\mathcal{F}$. For instance, $\{1\} \cup \{2\} = \{1,2\} \in \mathcal{F}$, and $\{1,3,5\} \cup \{2,4,6\} = \{1,2,3,4,5,6\} \in \mathcal{F}$.*

Exercise 1.2. *Let $A_1, A_2, \ldots, A_n$ be arbitrary subsets of Ω. Describe (explicitly) the smallest σ-algebra $\mathcal{F}$ containing $A_1, A_2, \ldots, A_n$. How many sets are there in $\mathcal{F}$? (Give an attainable upper bound under certain conditions). List all the sets in $\mathcal{F}$ for $n = 2$.*

Solution 1.2. ***Smallest σ-algebra containing $A_1, A_2, \ldots, A_n$:***

The smallest σ-algebra $\mathcal{F}$ containing the subsets $A_1, A_2, \ldots, A_n$ is generated by these sets. This means $\mathcal{F}$ includes all possible unions, intersections, and complements of these sets.

To explicitly describe $\mathcal{F}$:
1. Include $A_1, A_2, \ldots, A_n$.
2. Include the complements of each set: $A_1^c, A_2^c, \ldots, A_n^c$.
3. Include all possible unions and intersections of these sets and their complements.

Counting the Sets in $\mathcal{F}$:

In the worst-case scenario, if $A_1, A_2, \ldots, A_n$ are arbitrary subsets with no restrictions, the number of distinct sets that can be formed is determined by the combinations of unions and intersections. An attainable upper bound for the number of sets in $\mathcal{F}$ can be given by:

$$|\mathcal{F}| \leq 2^{2^n}$$

This upper bound arises from considering all subsets of Ω formed by the possible intersections of the $2n$ sets (including both original sets and their complements).

Example for $n = 2$:

Let A_1 and A_2 be two arbitrary subsets of Ω. The smallest σ-algebra $\mathcal{F}$ generated by A_1 and A_2 contains the following sets: A_1, A_2, A_1^c, A_2^c, $A_1 \cap A_2$, $A_1 \cap A_2^c$, $A_1^c \cap A_2$ and $A_1^c \cap A_2^c$.

Thus, the sets in $\mathcal{F}$ when $n = 2$ are:

$$\mathcal{F} = \{A_1, A_2, A_1^c, A_2^c, A_1 \cap A_2, A_1 \cap A_2^c, A_1^c \cap A_2, A_1^c \cap A_2^c\}$$

Exercise 1.3. *Let F and G be two σ-algebras of subsets of Ω.*
(a) *Is $F \cup G$, the collection of subsets of Ω lying in either F or G, a σ-algebra?*
(b) *Show that $F \cap G$, the collection of subsets of Ω lying in both F and G, is a σ-algebra.*
(c) *Generalize (b) to arbitrary intersections as follows. Let I be an arbitrary index set (possibly uncountable), and let $\{F_i\}_{i \in I}$ be a collection of σ-algebras on Ω. Show that $\bigcap_{i \in I} F_i$ is also a σ-algebra.*

Solution 1.3. *To determine whether $F \cup G$ is a σ-algebra, we need to check the three properties:*

Contains the empty set and sample space: *Since both F and G are σ-algebras, they each contain $\emptyset$ and Ω. Thus, $F \cup G$ contains both $\emptyset$ and Ω.*

Closed under complementation: *Let $A \in F \cup G$. If $A \in F$, then $A^c \in F$ (since F is a σ-algebra), and similarly for G. However, A^c might not be in $F \cup G$ if A is in one algebra but not in the other. Thus, $F \cup G$ is not closed under complementation.*

Closed under countable unions: *Let $A_1, A_2, \ldots \in F \cup G$. If all A_i are in F, then $\bigcup_{i=1}^{\infty} A_i \in F$. If all A_i are in G, then $\bigcup_{i=1}^{\infty} A_i \in G$. However, if some A_i are in F and some in G, $\bigcup_{i=1}^{\infty} A_i$ may not be in $F \cup G$. Therefore, $F \cup G$ is not closed under countable unions.*

*Hence, $F \cup G$ is **not a σ-algebra**.*

To show that $F \cap G$ is a σ-algebra, we verify the three properties:

Contains the empty set and sample space: *Since both F and G contain $\emptyset$ and Ω, we have $\emptyset \in F \cap G$ and $\Omega \in F \cap G$.*

Closed under complementation: *Let $A \in F \cap G$. Then $A \in F$ and $A \in G$. Thus, $A^c \in F$ and $A^c \in G$, which implies $A^c \in F \cap G$.*

Closed under countable unions: *Let $A_1, A_2, \ldots \in F \cap G$. Then $A_i \in F$ for all i and $A_i \in G$ for all i. Thus, $\bigcup_{i=1}^{\infty} A_i \in F$ and $\bigcup_{i=1}^{\infty} A_i \in G$, which implies $\bigcup_{i=1}^{\infty} A_i \in F \cap G$.*

*Therefore, $F \cap G$ **is a σ-algebra**.*

To prove that $\bigcap_{i \in I} F_i$ is a σ-algebra, we check the three properties:

Contains the empty set and sample space: *Since each F_i contains $\emptyset$ and Ω, we have $\emptyset \in \bigcap_{i \in I} F_i$ and $\Omega \in \bigcap_{i \in I} F_i$.*

Closed under complementation: *Let $A \in \bigcap_{i \in I} F_i$. Then $A \in F_i$ for all i. Thus, $A^c \in F_i$ for all i, which implies $A^c \in \bigcap_{i \in I} F_i$.*

Closed under countable unions: *Let $A_1, A_2, \ldots \in \bigcap_{i \in I} F_i$. Then $A_j \in F_i$ for all j and for all i. Thus, $\bigcup_{j=1}^{\infty} A_j \in F_i$ for all i, which implies $\bigcup_{j=1}^{\infty} A_j \in \bigcap_{i \in I} F_i$.*

*Therefore, $\bigcap_{i \in I} F_i$ **is a σ-algebra**.*

Exercise 1.4. *Let Ω be an arbitrary set. Answer the following questions:*
(a) *Is the collection F_1 consisting of all finite subsets of Ω an algebra?*
(b) *Let F_2 consist of all finite subsets of Ω and all subsets of Ω having a finite complement. Is F_2 an algebra?*
(c) *Is F_2 a σ-algebra?*
(d) *Let F_3 consist of all countable subsets of Ω and all subsets of Ω having a countable complement. Is F_3 a σ-algebra?*

Solution 1.4. *To determine if F_1 is an algebra, we must check the three properties:*

1. ***Contains the empty set:*** *$\emptyset \in F_1$ since the empty set is a finite subset.*
2. ***Closed under complementation:*** *If $A \in F_1$ (i.e., A is a finite subset of Ω), then its complement A^c may not be finite. Therefore, F_1 is not closed under complementation.*
3. ***Closed under finite unions:*** *If $A, B \in F_1$, then $A \cup B$ is also finite, so F_1 is closed under finite unions.*

*Since F_1 fails to be closed under complementation, we conclude that F_1 is **not an algebra**.*

To check if F_2 is an algebra, we verify the properties:

1. ***Contains the empty set:*** *$\emptyset \in F_2$ since it is a finite subset.*
2. ***Closed under complementation:***
   - *If $A \in F_2$ is finite, then A^c has a finite complement, which is infinite. Thus, it is in F_2.*
   - *If $B \in F_2$ has a finite complement, then B^c is finite. Therefore, $B^c \in F_2$.*

   Hence, F_2 is closed under complementation.
3. ***Closed under finite unions:***
   - *If $A, B \in F_2$ are both finite, then $A \cup B$ is finite.*
   - *If A is finite and B has a finite complement, then $A \cup B$ has a finite complement.*
   - *If both A and B have finite complements, then $(A \cup B)^c = A^c \cap B^c$, which is finite.*

Thus, F_2 is closed under finite unions.

To determine if F_2 is a σ-algebra, we need to check the closure under countable unions.

Consider the countable union of finite sets:

$$A_1 = \{1\}, A_2 = \{2\}, A_3 = \{3\}, \ldots$$

Then,

$$\bigcup_{i=1}^{\infty} A_i = \{1, 2, 3, \ldots\}$$

which is not finite. Therefore, F_2 is not closed under countable unions.

*Thus, F_2 is **not a σ-algebra**.*

To check if F_3 is a σ-algebra, we verify:

1. ***Contains the empty set:*** *$\emptyset \in F_3$ since it is countable.*

2. ***Closed under complementation:***

   - *If $A \in F_3$ is countable, then A^c has a countable complement.*
   - *If $B \in F_3$ has a countable complement, then B^c is countable.*

   Hence, F_3 is closed under complementation.

3. ***Closed under countable unions:***

   - *If $A_1, A_2, A_3, \ldots$ are countable sets, then*
   $$\bigcup_{i=1}^{\infty} A_i$$
   is also countable.
   - *If B has a countable complement, then*
   $$B^c \in F_3 \implies B^c = \bigcup_{i=1}^{\infty} C_i \quad \text{for } C_i \text{ countable.}$$
   Therefore, B itself is in F_3.

*Since F_3 satisfies all properties, we conclude that F_3 is **a σ-algebra**.*

Exercise 1.5. *Let X and Y be two sets and let $f : X \to Y$ be a function. If F is a σ-algebra over the subsets of Y, and $G = \{A \mid \exists B \in F \text{ such that } f^{-1}(B) = A\}$, does G form a σ-algebra of subsets of X? Note that $f^{-1}(N)$ is the notation used for the pre-image of set N under the function f for some $N \subseteq Y$. That is, $f^{-1}(N) = \{x \in X \mid f(x) \in N\}$ for some $N \subseteq Y$.*

Solution 1.5. *To show that G forms a σ-algebra of subsets of X, we need to verify that G satisfies the three properties of a σ-algebra:*

Contains the empty set*: The σ-algebra F over Y contains the empty set, $\emptyset$. Let $B = \emptyset \in F$. Then the pre-image under f, $f^{-1}(\emptyset) = \emptyset$, is also in G. Therefore, $\emptyset \in G$.*

Closed under complementation*: Let $A \in G$. By definition of G, there exists a set $B \in F$ such that $f^{-1}(B) = A$. We need to show that $A^c \in G$. Consider the complement of B in F, denoted as B^c. Since F is a σ-algebra, $B^c \in F$.*

Now, observe that:
$$f^{-1}(B^c) = \{x \in X \mid f(x) \notin B\} = A^c$$
Hence, $A^c \in G$, showing that G is closed under complementation.

Closed under countable unions*: Let $\{A_i\}_{i=1}^{\infty}$ be a countable collection of sets in G. For each $A_i \in G$, there exists $B_i \in F$ such that $f^{-1}(B_i) = A_i$. Since F is a σ-algebra, it is closed under countable unions, so $\bigcup_{i=1}^{\infty} B_i \in F$.*

Now, consider the pre-image:

$$f^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(B_i) = \bigcup_{i=1}^{\infty} A_i$$

Therefore, $\bigcup_{i=1}^{\infty} A_i \in G$, showing that G is closed under countable unions.

1.2 Measurable Space

Consider a large canvas Ω, which represents an entire art studio wall. The goal is to measure the total amount of paint used on specific regions of the wall. However, you don't want to measure paint usage for every possible shape or region on the wall (which could be infinitely complex). Instead, you decide to focus only on certain, manageable regions such as: rectangles, circles, simple polygons, and unions and intersections of these shapes.

These regions form a collection, say $\mathcal{F}$, which represents all the shapes and combinations that you are interested in measuring. The pair $(\Omega, \mathcal{F})$ then becomes a *measurable space*, where:

- Ω is the entire canvas, representing all possible points on the wall.
- $\mathcal{F}$ is a collection of specific shapes (rectangles, circles, etc.) and their combinations, which are the regions you can measure the amount of paint for.

The measurable sets in $\mathcal{F}$ are those specific shapes and combinations that you have chosen to focus on, similar to how measurable sets in probability theory are those events that belong to a specific σ-algebra. In context of probability theory, imagine you have a sample space Ω, which represents all the possible outcomes of an experiment. For instance, if you flip a coin, the sample space is $\Omega = \{\text{Heads}, \text{Tails}\}$.

A measurable space is a pair $(\Omega, \mathcal{F})$, where:

- Ω is the sample space, representing all possible outcomes.
- $\mathcal{F}$ is a σ-algebra of subsets of Ω. It is a collection of subsets that includes the empty set, is closed under complements, and closed under countable unions.

The subsets in $\mathcal{F}$ are the ones we can *measure*, hence the term *measurable space*.

Definition 1.3. *The 2-tuple $(\Omega, \mathcal{F})$ is called a* ***measurable space****. Here:*

- *Ω is the sample space, a non-empty set.*
- *$\mathcal{F}$ is a σ-algebra on Ω, meaning it satisfies:*
    1. *$\emptyset \in \mathcal{F}$.*
    2. *If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ (closure under complements).*
    3. *If $A_1, A_2, A_3, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (closure under countable unions).*

Definition 1.4. *Every member of the σ-algebra $\mathcal{F}$ is called an $\mathcal{F}$**-measurable set** in the context of measure theory. These are the subsets of Ω that we can measure using the σ-algebra $\mathcal{F}$.*

Definition 1.5. *$\mathcal{F}$-measurable sets are called **events**. This means that an event is not just any subset of Ω, but one that belongs to the σ-algebra $\mathcal{F}$ under consideration.*

Examples Where Subsets of Ω Are Not $\mathcal{F}$-Measurable Sets:

In measure theory and probability, not all subsets of a sample space Ω are necessarily $\mathcal{F}$-measurable sets. This depends on the construction of the σ-algebra $\mathcal{F}$ associated with Ω. Below are some examples where subsets of Ω are not $\mathcal{F}$-measurable:

Example 1: A Countable Union Not in an Algebra

Let $\Omega = \mathbb{N}$, the set of natural numbers, and let $\mathcal{F}$ be an **algebra** consisting of all finite subsets of $\mathbb{N}$ and their complements (which are cofinite sets). In this setup, $\mathcal{F}$ contains only finite unions and intersections.

Now, consider the subset $A = \{2, 4, 6, \ldots\}$, the set of all even numbers. This is an **infinite** set but not cofinite (its complement, the set of odd numbers, is also infinite). Since $\mathcal{F}$ only contains finite or cofinite sets, A is not in $\mathcal{F}$. Thus, A is an example of a subset of Ω that is not $\mathcal{F}$-measurable.

Example 2: Subsets in the Cantor Set

Let Ω be the **Cantor set**, which is a subset of the interval $[0, 1]$. Construct $\mathcal{F}$ to be the σ-algebra generated by all **intervals** in $[0, 1]$. While $\mathcal{F}$ will contain many subsets, it will not include certain highly irregular subsets of the Cantor set that are not expressible as a countable union, intersection, or complement of intervals.

For instance, a subset of the Cantor set that is formed using a complex pattern based on the binary expansion of its elements may not be measurable in this σ-algebra. Hence, such a subset would not be $\mathcal{F}$-measurable.

Example 3: Students in a Class

Imagine a class of 30 students, represented by the set:

$$\Omega = \{s_1, s_2, s_3, \ldots, s_{30}\}$$

Define a σ-algebra $\mathcal{F}$ that consists only of subsets containing an even number of students. This σ-algebra could include sets like:

$$\mathcal{F} = \{\emptyset, \{s_1, s_2\}, \{s_3, s_4\}, \{s_5, s_6\}, \ldots, \{s_1, s_2, s_3, s_4\}, \ldots, \{s_1, s_2, \ldots, s_{30}\}\}$$

Now, consider the subset of interest $A = \{s_1, s_3, s_5, s_7, \ldots, s_{29}\}$, which contains all the odd-numbered students in the class.

In the σ-algebra $\mathcal{F}$, every set is constructed to contain only an even number of students. The set A, which contains an odd number of students, cannot be expressed as a union or intersection of sets from $\mathcal{F}$.

Thus, A is not $\mathcal{F}$-measurable because it does not fit within the constraints of our σ-algebra.

Example 4: Days of Week

Let $\Omega = \{\text{Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday}\}$. Suppose we define a σ-algebra $\mathcal{F}$ that only includes subsets that contain weekdays:

$$\mathcal{F} = \{\emptyset, \{\text{Monday, Tuesday, Wednesday, Thursday, Friday}\}, \{\text{Saturday, Sunday}\}, \Omega\}$$

Now, consider the subset of interest $A = \{\text{Saturday}\}$.

The σ-algebra $\mathcal{F}$ only contains the sets of weekdays and their complements but does not include individual weekend days like Saturday. Thus, A cannot be constructed as a union or intersection of sets in $\mathcal{F}$.

Since A cannot be represented within the existing σ-algebra $\mathcal{F}$, it is not $\mathcal{F}$-measurable.

1.2.1 Measure

Definition 1.6. *Let $(\Omega, \mathcal{F})$ be a measurable space. A measure on $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ such that:*

1. $\mu(\emptyset) = 0$.
2. *If $\{A_i, i \geq 1\}$ is a sequence of disjoint sets in $\mathcal{F}$, then the measure of the union of these countably infinite disjoint sets is equal to the sum of the measures of the individual sets:*

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

The second property stated above is known as the *countable additivity property* of measures. From the definition, it is clear that a measure can only be assigned to elements of $\mathcal{F}$.

Definition 1.7. *The triplet $(\Omega, \mathcal{F}, \mu)$ is called a measure space.*

The measure μ is said to be a *finite measure* if $\mu(\Omega) < \infty$; otherwise, μ is said to be an *infinite measure*. In particular, if $\mu(\Omega) = 1$, then μ is referred to as a *probability measure*.

1.2.2 Probability Measure

Definition 1.8. *A probability measure is a function $P : \mathcal{F} \to [0, 1]$ such that:*

1. $P(\emptyset) = 0$.
2. $P(\Omega) = 1$.

3. ***Countable Additivity:*** *If $\{A_i, i \geq 1\}$ is a sequence of disjoint sets in $\mathcal{F}$, then*

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

Definition 1.9. *The triplet $(\Omega, \mathcal{F}, P)$ is called a probability space, and the three properties stated above are referred to as the axioms of probability.*

It is clear from the definition that probabilities are defined only for elements of $\mathcal{F}$, and not necessarily for all subsets of Ω. In other words, probability measures are assigned only to *events*. Even when we speak of the probability of an elementary outcome ω, it should be interpreted as the probability assigned to the singleton set $\{\omega\}$ (assuming, of course, that the singleton is an event).

1.2.3 Properties of Probability Measure

We will derive some fundamental properties of probability measures, which follow directly from the axioms of probability. In what follows, $(\Omega, \mathcal{F}, P)$ is a probability space.

Property 1: Suppose A be a subset of Ω such that $A \in \mathcal{F}$. Then,

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

Proof: Given any subset $A \subseteq \Omega$, A and A^c partition the sample space. Hence, $A^c \cup A = \Omega$ and $A^c \cap A = \emptyset$. By the "Countable Additivity" axiom of probability, $P(A^c \cup A) = P(A) + P(A^c)$. Therefore, $P(\Omega) = P(A) + P(A^c) \implies P(A^c) = 1 - P(A)$.

Property 2: Consider events A and B such that $A \subseteq B$ and $A, B \in \mathcal{F}$. Then $P(A) \leq P(B)$.

Proof: The set B can be written as the union of two disjoint sets A and $A^c \cap B$. Therefore, we have $P(A) + P(A^c \cap B) = P(B) \implies P(A) \leq P(B)$ since $P(A^c \cap B) \geq 0$.

Property 3: (Finite Additivity) If $A_1, A_2, \ldots, A_n$ are a finite number of disjoint events, then

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

Proof: This property follows directly from the axiom of countable additivity of probability measures. It is obtained by setting the events $A_{n+1}, A_{n+2}, \ldots$ as empty sets. The left-hand side (LHS) will simplify as:

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

The right-hand side (RHS) can be manipulated as follows:

$$\sum_{i=1}^{\infty} P(A_i) \stackrel{(a)}{=} \lim_{k\to\infty} \sum_{i=1}^{k} P(A_i) = \sum_{i=1}^{n} P(A_i) + \lim_{k\to\infty} \sum_{i=n+1}^{k} P(A_i) \stackrel{(b)}{=} \sum_{i=1}^{n} P(A_i) + \lim_{k\to\infty} 0 = \sum_{i=1}^{n} P(A_i).$$

where (a) follows from the definition of an infinite series and (b) is a consequence of setting the events from A_{n+1} onwards to null sets.

Property 4: For any $A, B \in F$,

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

In general, for a family of events $\{A_i\}_{i=1}^{n} \subset F$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right).$$

This property is proved using induction on n. The property can be proved in a much simpler way using the concept of Indicator Random Variables, which will be discussed in the subsequent lectures.

Proof The set $A \cup B$ can be written as $A \cup B = A \cup (A^c \cap B)$. Since A and $A^c \cap B$ are disjoint events, $P(A \cup B) = P(A) + P(A^c \cap B)$. Now, set B can be partitioned as $B = (A \cap B) \cup (A^c \cap B)$. Hence, $P(B) = P(A \cap B) + P(A^c \cap B)$. On substituting this result in the expression of $P(A \cup B)$, we will obtain the final result that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Property 5: If $\{A_i, i \geq 1\}$ are events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{m \to \infty} P\left(\bigcup_{i=1}^{m} A_i\right).$$

This result is known as the continuity of probability measures.

How to visualise this property? Imagine P_m is the probability of union of $A_1, A_2, ..., A_m$. Then the sequence of P_m's is a monotonically increasing sequence. Also, the sequence is bounded by the interval $[0, 1]$. We know, that every monotonically increasing sequence that is bounded must converge. So the RHS of the property 5 is a finite quantity. So is the LHS because the countable union of sets is a well-definied set for which the probability measure is defined. The property says both are equal.

Proof: Define a new family of sets $B_1 = A_1, B_2 = A_2 \setminus A_1, \ldots, B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i, \ldots$.

Then, the following claims are placed:
Claim 1: $B_i \cap B_j = \emptyset, \forall i \neq j$.
Claim 2: $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$.

Since $\{B_i, i \geq 1\}$ is a disjoint sequence of events, and using the above claims, we get

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

Therefore,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(B_i) \quad (a) = \lim_{m \to \infty} \sum_{i=1}^{m} P(B_i) \quad (b) = \lim_{m \to \infty} P\left(\bigcup_{i=1}^{m} B_i\right)$$

$$(c) = \lim_{m\to\infty} P\left(\bigcup_{i=1}^{m} A_i\right).$$

Here, (a) follows from the definition of an infinite series, (b) follows from Claim 1 in conjunction with the Countable Additivity axiom of probability measure, and (c) follows from the intermediate result required to prove Claim 2. Hence proved.

Property 6: If $\{A_i, i \geq 1\}$ is a sequence of increasing nested events i.e., $A_i \subseteq A_{i+1}, \forall i \geq 1$, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{m\to\infty} P(A_m).$$

Property 7: If $\{A_i, i \geq 1\}$ is a sequence of decreasing nested events i.e., $A_{i+1} \subseteq A_i, \forall i \geq 1$, then

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{m\to\infty} P(A_m).$$

Properties 6 and 7 are said to be corollaries to Property 5.

Property 8: Suppose $\{A_i, i \geq 1\}$ are events, then

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

This result is known as the Union Bound. This bound is trivial if $\sum_{i=1}^{\infty} P(A_i) \geq 1$ since the LHS of Property 8 is a probability of some event. This is a very widely used bound, and has several applications. For instance, the union bound is used in the probability of error analysis in Digital Communications for complicated modulation schemes.

Proof: Define a new family of sets $B_1 = A_1, B_2 = A_2 \setminus A_1, \ldots, B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i, \ldots$.

Claim 1: $B_i \cap B_j = \emptyset, \forall i \neq j$.
Claim 2: $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$.

Since $\{B_i, i \geq 1\}$ is a disjoint sequence of events, and using the above claims, we get

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

Also, since $B_i \subseteq A_i \, \forall i \geq 1$, $P(B_i) \leq P(A_i) \, \forall i \geq 1$ (using Property 2). Therefore, the finite sum of probabilities follows

$$\sum_{i=1}^{n} P(B_i) \leq \sum_{i=1}^{n} P(A_i).$$

Eventually, in the limit, the following holds:

$$\sum_{i=1}^{\infty} P(B_i) \leq \sum_{i=1}^{\infty} P(A_i).$$

Finally, we arrive at the result,

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

Exercise 1.6. *A standard card deck (52 cards) is distributed to two persons: 26 cards to each person. All partitions are equally likely. Find the probability that the first person receives all four aces.*

Solution 1.6. *To find the probability that the first person receives all four aces when a standard deck of 52 cards is distributed equally between two persons (each receiving 26 cards), we use a measure-theoretic approach.*

Let $(\Omega, \mathcal{F}, P)$ be the probability space where:
- Ω is the set of all ways to partition the deck into two hands of 26 cards each.
- $\mathcal{F}$ is the σ-algebra of subsets of Ω.
- P is the uniform probability measure on $(\Omega, \mathcal{F})$.

Step 1: Total Number of Outcomes

The total number of ways to choose 26 cards out of 52 for the first person is given by:

$$|\Omega| = \binom{52}{26}$$

where $\binom{52}{26}$ denotes the binomial coefficient representing the number of ways to choose 26 cards from 52.

Step 2: Number of Favorable Outcomes

Next, we find the number of ways the first person can receive all four aces. If the first person is to receive all four aces, we must choose the remaining 22 cards from the remaining 48 non-ace cards. The number of ways to do this is:

$$|\Omega_{favorable}| = \binom{48}{22}$$

where $\binom{48}{22}$ denotes the binomial coefficient representing the number of ways to choose 22 cards from the 48 non-ace cards.

Step 3: Calculating the Probability

The probability that the first person receives all four aces is the ratio of the number of favorable outcomes to the total number of outcomes:

$$P(\textit{First person receives all four aces}) = \frac{|\Omega_{favorable}|}{|\Omega|}$$

$$P(\textit{First person receives all four aces}) = \frac{\binom{48}{22}}{\binom{52}{26}}$$

Exercise 1.7. *Let $\{A_r\}_{r=1}^n$ be a finite collection of events in a probability space $(\Omega, \mathcal{F}, P)$. We aim to prove that:*

$$P\left(\bigcup_{1\leq r\leq n} A_r\right) \leq \min_{1\leq k\leq n}\left\{\sum_{1\leq r\leq n} P(A_r) - \sum_{r:r\neq k} P(A_r\cap A_k)\right\}$$

Solution 1.7. *Define $S = \bigcup_{r=1}^n A_r$. By the inclusion-exclusion principle for a finite union of events, we have:*

$$P(S) = \sum_{1\leq r\leq n} P(A_r) - \sum_{1\leq r<s\leq n} P(A_r\cap A_s) + \ldots + (-1)^{n+1}P\left(\bigcap_{1\leq r\leq n} A_r\right).$$

This expression accounts for all possible intersections of the events A_r. However, to prove the inequality, we'll make use of the following upper bound:

Consider any fixed $k \in \{1, 2, \ldots, n\}$. We can express $P(S)$ as:

$$P(S) \leq P(A_k) + P\left(\bigcup_{r:r\neq k} A_r \setminus A_k\right).$$

This follows since the probability of S is at most the probability of A_k plus the probability of the events outside of A_k but not overlapping with it.

Now, observe that:

$$P\left(\bigcup_{r:r\neq k} A_r \setminus A_k\right) \leq \sum_{r:r\neq k} P(A_r\setminus A_k).$$

Using the identity $P(A_r\setminus A_k) = P(A_r) - P(A_r\cap A_k)$, we can rewrite the above as:

$$P\left(\bigcup_{r:r\neq k} A_r \setminus A_k\right) \leq \sum_{r:r\neq k} \left(P(A_r) - P(A_r\cap A_k)\right).$$

Therefore:

$$P(S) \leq P(A_k) + \sum_{r:r\neq k} \left(P(A_r) - P(A_r\cap A_k)\right).$$

Simplifying further:

$$P(S) \leq \sum_{1\leq r\leq n} P(A_r) - \sum_{r:r\neq k} P(A_r\cap A_k).$$

Since this inequality holds for any $k \in \{1, 2, \ldots, n\}$, we take the minimum over all k:

$$P\left(\bigcup_{1\leq r\leq n} A_r\right) \leq \min_{1\leq k\leq n}\left\{\sum_{1\leq r\leq n} P(A_r) - \sum_{r:r\neq k} P(A_r\cap A_k)\right\}.$$

Exercise 1.8. *You are given that at least one of the events A_n, $1 \leq n \leq N$, is certain to occur. However, certainly no more than two occur. If $P(A_n) = p$ and $P(A_n \cap A_m) = q$, $m \neq n$, then show that $p \geq \frac{1}{N}$ and $q \leq \frac{2}{N}$.*

Solution 1.8. *Given the events $A_1, A_2, \ldots, A_N$ such that at least one event occurs and at most two occur, we have:*

$$P\left(\bigcup_{n=1}^{N} A_n\right) = 1$$

and

$$P\left(A_n \cap A_m\right) = q \quad \textit{for} \quad m \neq n.$$

By the principle of inclusion-exclusion, the probability of the union of these events is:

$$P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{n=1}^{N} P(A_n) - \sum_{1 \leq n < m \leq N} P(A_n \cap A_m).$$

Substituting the given values, we get:

$$1 = \sum_{n=1}^{N} p - \sum_{1 \leq n < m \leq N} q.$$

The number of terms in the first sum is N, so:

$$1 = Np - \binom{N}{2} q,$$

where $\binom{N}{2} = \frac{N(N-1)}{2}$ is the number of ways to choose 2 events from N.

Thus:

$$1 = Np - \frac{N(N-1)}{2} q.$$

Rearranging to solve for p:

$$Np = 1 + \frac{N(N-1)}{2} q,$$

$$p = \frac{1}{N} + \frac{(N-1)}{2} q.$$

Since at most two events can occur, q must be small enough so that no three events can occur simultaneously. Hence, by substituting $p \geq \frac{1}{N}$:

$$\frac{1}{N} + \frac{(N-1)}{2} q \geq \frac{1}{N},$$

$$\frac{(N-1)}{2} q \geq 0.$$

To find the upper bound of q, note that since at most two events can occur, the total probability contributed by the intersections should not exceed 1. Therefore:

$$\frac{(N-1)}{2} q \leq \frac{1}{N}.$$

$$q \leq \frac{2}{N}.$$

Exercise 1.9. *Consider a measurable space $(\Omega, \mathcal{F})$ with $\Omega = [0,1]$. A measure P is defined on the non-empty subsets of Ω (in $\mathcal{F}$), which are all of the form (a,b), $(a,b]$, $[a,b)$ and $[a,b]$, as the length of the interval, i.e.,*

$$P((a,b)) = P((a,b]) = P([a,b)) = P([a,b]) = b - a.$$

(a) *Show that P is not just a measure, but it's a probability measure.*
(b) *Let $A_n = \left[\frac{1}{n+1}, 1\right]$ and $B_n = \left[0, \frac{1}{n+1}\right]$ for $n \geq 1$. Compute $P\left(\cup_{i\in\mathbb{N}} A_i\right)$, $P\left(\cap_{i\in\mathbb{N}} A_i\right)$, $P\left(\cup_{i\in\mathbb{N}} B_i\right)$, and $P\left(\cap_{i\in\mathbb{N}} B_i\right)$.*
(c) *Compute $P\left(\cap_{i\in\mathbb{N}} (B_i^c \cup A_i^c)\right)$.*
(d) *Let $C_m = \left[0, \frac{1}{m}\right]$ such that $P(C_m) = P(A_n)$. Express m in terms of n.*
(e) *Evaluate $P\left(\cap_{i\in\mathbb{N}} (C_i \cap A_i)\right)$ and $P\left(\cup_{i\in\mathbb{N}} (C_i \cap A_i)\right)$.*

Solution 1.9. *To show that P is a probability measure, we need to verify two properties:*

1. Non-negativity: $P(A) \geq 0$ for all $A \in \mathcal{F}$. By definition, $P(A) = b - a \geq 0$ since $b \geq a$ for all intervals in $[0,1]$.

2. $P(\Omega) = 1$: The entire space $\Omega = [0,1]$. Hence, $P([0,1]) = 1 - 0 = 1$.

Therefore, P is a probability measure.

$P\left(\cup_{i\in\mathbb{N}} A_i\right) = P([0,1]) = 1$
$P\left(\cap_{i\in\mathbb{N}} A_i\right) = P\left([0,1]\right) = 1$
$P\left(\cup_{i\in\mathbb{N}} B_i\right) = P([0,1]) = 1$
$P\left(\cap_{i\in\mathbb{N}} B_i\right) = P(\{0\}) = 0$

Note that $B_i^c = [\frac{1}{n+1}, 1]$ and $A_i^c = [0, \frac{1}{n+1}]$. Thus:

$$P\left(\cap_{i\in\mathbb{N}} (B_i^c \cup A_i^c)\right) = P(\emptyset) = 0$$

We have $P(C_m) = \frac{1}{m}$ and $P(A_n) = 1 - \frac{1}{n+1}$. Equating these gives:

$$\frac{1}{m} = 1 - \frac{1}{n+1}$$

$$m = \frac{n+1}{n}$$

$P\left(\cap_{i\in\mathbb{N}} (C_i \cap A_i)\right) = P(\emptyset) = 0$
$P\left(\cup_{i\in\mathbb{N}} (C_i \cap A_i)\right) = P([0,1]) = 1$

1.3 Discrete Probability Spaces

Discrete probability spaces correspond to the case when the sample space Ω is countable. This is the most conceptually straightforward case, since it is possible to assign probabilities to all subsets of Ω.

Definition 1.10. *A probability space $(\Omega, \mathcal{F}, P)$ is said to be a discrete probability space if the following conditions hold:*

(a) The sample space Ω is finite or countably infinite,

(b) The σ-algebra is the set of all subsets of Ω, i.e., $\mathcal{F} = 2^{\Omega}$, and

(c) The probability measure, P, is defined for every subset of Ω. In particular, it can be defined in terms of the probabilities $P(\{\omega\})$ of the singletons corresponding to each of the elementary outcomes ω, and satisfies for every $A \in \mathcal{F}$,

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

and

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

The above definition highlights that it is possible to assign probabilities to each singleton set of Ω, but it doesn't say about *what probabilities to assign?* This depends on our use-case and what we want to model.

Examples of Discrete Probability Spaces

1. Let us consider a coin toss experiment with the probability of getting a head as p and the probability of getting a tail as $(1-p)$. The sample space and the σ-algebra are defined as follows:

$$\Omega = \{H, T\} \equiv \{0, 1\}, \quad \mathcal{F} = 2^{\Omega} = \{\emptyset, \{H\}, \{T\}, \Omega\}.$$

The probability measure is given by:

$$P(\{H\}) \equiv P(\{0\}) = p, \quad P(\{T\}) \equiv P(\{1\}) = 1 - p.$$

In this case, we say that $P(.)$ is a Bernoulli measure on $\{\{0,1\}, 2^{\{0,1\}}\}$.

2. Let $\Omega = \mathbb{N}$ and $\mathcal{F} = 2^{\mathbb{N}}$. We can define the probability of a singleton as:

$$P(\{k\}) = a_k \geq 0, \quad k \in \mathbb{N},$$

under the constraint that:

$$\sum_{k \in \mathbb{N}} P(\{k\}) = 1.$$

For example, if we let $a_k = \frac{1}{2^k}$, $k \in \mathbb{N}$, this is a valid measure, since:

$$\sum_{k \in \mathbb{N}} \frac{1}{2^k} = 1.$$

As another example, consider $a_k = (1-p)^{k-1} p$ for $0 < p < 1$ and $k \in \mathbb{N}$. This is known as a geometric measure with parameter p. It is a valid probability measure since:

$$\sum_{k \in \mathbb{N}} (1-p)^{k-1} p = 1.$$

3. Let $\Omega = \mathbb{N} \cup \{0\}$ and $\mathcal{F} = 2^\Omega$. We define the probability measure as:

$$P(\{k\}) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \lambda > 0.$$

This probability measure is called a Poisson measure with parameter λ on $\{\Omega, 2^\Omega\}$. This is a valid probability measure, since:

$$\sum_{k=0}^{\infty} P(\{k\}) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

4. Let $\Omega = \{0, 1, 2, \ldots, N\}$, where $N \in \mathbb{N}$ and $\mathcal{F} = 2^\Omega$. We define the probability measure as:

$$P(\{k\}) = \binom{N}{k} p^k (1-p)^{N-k}, \quad 0 < p < 1.$$

This probability measure is called a Binomial measure with parameters (N, p) on $\{\Omega, 2^\Omega\}$. This can be verified to be a valid probability measure as follows:

$$\sum_{k \in \Omega} \binom{N}{k} p^k (1-p)^{N-k} = (p + (1-p))^N = 1.$$

Note that in all the examples above, we have not explicitly specified an expression for $P(A)$ for every $A \subset \Omega$. Since the sample space is countable, the probability of any subset of the sample space can be obtained as the sum of probabilities of the corresponding elementary outcomes. In other words, for discrete probability spaces, it suffices to specify the probabilities of singletons corresponding to each of the elementary outcomes.

Chapter 2

Borel Sets and Lebesgue Measure

Last chapter lays the foundation of *what are events* and *what is a probability measure*. But that still doesn't answer the question - *what can we actually measure and what can we not? And how to measure what we can measure?* Sounds tricky? Don't worry! This chapter will take you through the complications!

2.1 Introduction to Borel Sets

Let's consider the case when the sample space is uncountable.

2.1.1 Uncountable Sample Space

Consider the experiment of picking a real number at random from $\Omega = [0, 1]$, such that every number is *equally likely* to be picked. It is quite apparent that a simple strategy of assigning probabilities to singleton subsets of the sample space gets into difficulties quite quickly. Indeed:

(i) If we assign some positive probability to each elementary outcome, then the probability of an event with infinitely many elements, such as $A = \left\{1, \frac{1}{2}, \frac{1}{3}, \dots\right\}$, would become unbounded. This is because the sum of positive probabilities over an infinite set would diverge.

(ii) If we assign zero probability to each elementary outcome, this alone would not be sufficient to determine the probability of an uncountable subset of Ω, such as $\left[\frac{1}{2}, \frac{2}{3}\right]$. This is because probability measures are not additive over uncountable disjoint unions (of singletons in this case). Assigning zero probability to singletons does not directly imply how to handle intervals or other uncountable sets.

Thus, we need a different approach to assign probabilities when the sample space is uncountable, such as $\Omega = [0, 1]$. In particular, we need to assign probabilities directly to specific subsets of Ω. Intuitively, we would like our *uniform measure* μ on $[0, 1]$ to possess the following two properties:

(i) $\mu((a, b)) = \mu((a, b]) = \mu([a, b)) = \mu([a, b])$ for any interval in $[0, 1]$. This ensures that the measure is consistent across different types of intervals, capturing the idea of *equal likelihood* for any interval of the same length.

(ii) **Translational Invariance**: That is, if $A \subseteq [0,1]$, then for any $x \in \Omega$, $\mu(A \oplus x) = \mu(A)$, where the set $A \oplus x$ is defined as:

$$A \oplus x = \{a + x \mid a \in A, a + x \leq 1\} \cup \{a + x - 1 \mid a \in A, a + x > 1\}$$

This property ensures that the measure is invariant under translation within the interval $[0,1]$, reflecting the uniformity of the measure.

However, the following impossibility result asserts that there is no way to consistently define a uniform measure on all subsets of $[0,1]$. This result is rooted in the fact that certain sets in $[0,1]$ (those that are non-measurable) defy any consistent assignment of measure while preserving the desired properties of translation invariance and interval consistency.

Theorem 2.1. ***Impossibility Result:*** *There does not exist a definition of a measure $\mu(A)$ for all subsets of $[0,1]$ satisfying:*

(i) $\mu((a,b)) = \mu((a,b]) = \mu([a,b)) = \mu([a,b])$

(ii) Translational Invariance: If $A \subseteq [0,1]$, then for any $x \in \Omega$, $\mu(A \oplus x) = \mu(A)$, where the set $A \oplus x$ is defined as:

$$A \oplus x = \{a + x \mid a \in A, a + x \leq 1\} \cup \{a + x - 1 \mid a \in A, a + x > 1\}$$

Proof. To show the impossibility, we argue that a measure satisfying these two properties for all subsets of $[0,1]$ leads to a contradiction. We will use basic properties of measures and simple logic to establish the proof.

1. Interval Length Property (i):

The first property tells us that the measure of any interval in $[0,1]$ is simply the length of the interval. For example, if $A = (a,b)$, then $\mu((a,b)) = b - a$. This property holds for open, closed, and half-open intervals.

2. Translational Invariance (ii):

The second property states that if we shift a set A by some amount x, its measure should remain the same. For example, if A is an interval, shifting it within $[0,1]$ should not change its length. This makes sense intuitively, as the measure should not depend on the location of the set but only its size.

3. Partitioning $[0,1]$ into Equal Parts:

Let's divide the interval $[0,1]$ into n equal parts. Define sets $A_i = \left[\frac{i-1}{n}, \frac{i}{n}\right)$ for $i = 1, 2, \ldots, n-1$ and $A_n = \left[\frac{n-1}{n}, 1\right]$. By property (i), each of these sets has a measure:

$$\mu(A_i) = \frac{1}{n}, \quad \text{for all } i = 1, 2, \ldots, n.$$

Since these intervals are disjoint and together cover $[0, 1]$, by the additivity of measures:

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i) = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

4. Constructing Translations:

Suppose we take one of these intervals, say $A_1 = \left[0, \frac{1}{n}\right)$, and shift it by $\frac{1}{2n}$. This new set, $A_1 \oplus \frac{1}{2n}$, becomes $\left[\frac{1}{2n}, \frac{3}{2n}\right)$, which is a valid subset of $[0, 1]$.

By translational invariance (property (ii)), $\mu(A_1 \oplus \frac{1}{2n}) = \mu(A_1) = \frac{1}{n}$.

5. Forming a Contradiction:

Now, let's consider translating A_1 by different multiples of $\frac{1}{2n}$. We form the following sets:

$$A_1, \quad A_1 \oplus \frac{1}{2n}, \quad A_1 \oplus \frac{2}{2n}, \quad \ldots, \quad A_1 \oplus \frac{n-1}{2n}.$$

These n translations of A_1 are all disjoint, and by property (ii), each has a measure of $\frac{1}{n}$.

However, if we sum up the measures of all these disjoint translations, we get:

$$\mu(A_1) + \mu(A_1 \oplus \frac{1}{2n}) + \ldots + \mu(A_1 \oplus \frac{n-1}{2n}) = \frac{1}{n} \times n = 1.$$

6. Contradiction with the Total Measure:

Observe that the union of all these translated sets may not cover the entire interval $[0, 1]$. In fact, since A_1 is just a small fraction of $[0, 1]$, these translations form only a portion of $[0, 1]$. Hence, the measure of their union should be less than 1.

But by translational invariance and additivity, we have shown that the sum of their measures equals 1. This creates a contradiction because it implies that a part of $[0, 1]$ has the same measure as the whole interval.

Since this contradiction arises when attempting to define μ on all subsets while preserving both the interval property and translational invariance, it is impossible to define such a measure for all subsets of $[0, 1]$. □

Therefore, we must compromise, and consider a smaller σ-algebra that contains certain *nice* subsets of the sample space [0, 1]. **These *nice* subsets are the intervals**, and the resulting σ-algebra is called the Borel σ-algebra.

Before defining Borel sets, we introduce the concept of generating σ-algebras from a given collection of subsets.

2.1.2 Borel Sets

Now, we know that the collection of intervals is not a σ-algebra because if $[a, b]$ is in the collection, it's complement is not an interval. So, we want to build towards a σ-algebra that contains all intervals, their complements and is closed under countable unions and countable intersections.

Let $\mathcal{C}$ be the collection of all nice subsets of sample space Ω in which we are interested. We have to generate the smallest σ-algebra that contains $\mathcal{C}$, that is denoted by $\sigma(\mathcal{C})$.

Theorem 2.2. *The intersection of an arbitrary number of σ-algebras is a σ-algebra.*

Proof. Let $\{\mathcal{F}_i\}_{i\in I}$ be a collection of σ-algebras on a collection $\mathcal{C}$, where I is an index set. Define $\mathcal{F} = \bigcap_{i\in I} \mathcal{F}_i$. We want to show that $\mathcal{F}$ is a σ-algebra.

Since each $\mathcal{F}_i$ is a σ-algebra, it contains $\mathcal{C}$. Therefore, $\mathcal{C} \in \mathcal{F}_i$ for all $i \in I$. By definition of the intersection, $\mathcal{C} \in \mathcal{F}$.

Let $A \in \mathcal{F}$. This implies that $A \in \mathcal{F}_i$ for every $i \in I$. Since each $\mathcal{F}_i$ is a σ-algebra, $A^c \in \mathcal{F}_i$ for all $i \in I$. Therefore, $A^c \in \bigcap_{i\in I} \mathcal{F}_i = \mathcal{F}$. Hence, $\mathcal{F}$ is closed under complementation.

Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of sets in $\mathcal{F}$. This implies that $A_n \in \mathcal{F}_i$ for every $i \in I$ and for all $n \in \mathbb{N}$. Since each $\mathcal{F}_i$ is a σ-algebra, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_i$ for all $i \in I$. Therefore, $\bigcup_{n=1}^{\infty} A_n \in \bigcap_{i\in I} \mathcal{F}_i = \mathcal{F}$. Hence, $\mathcal{F}$ is closed under countable unions.

Since $\mathcal{F}$ contains $\mathcal{C}$, is closed under complementation, and is closed under countable unions, $\mathcal{F}$ is a σ-algebra.

□

Theorem 2.3. *The smallest σ-algebra, $\sigma(\mathcal{C})$ is $\mathcal{F} = \bigcap_{i\in I} \mathcal{F}_i$, where each $\mathcal{F}_i$ is the σ-algebra of $\mathcal{C}$.*

The above theorem just implies that the smallest σ-algebra exists and is well defined. We don't know all $\mathcal{F}_i$, and we don't intend to find them as well. What we know for now is - that they exist. And if we take all of them and take a countable intersection of them - the resultant collection of sets is well-defined.

Proof. Let $\{\mathcal{F}_i\}_{i\in I}$ be the collection of all σ-algebras on a set $\mathcal{C}$ that contain $\mathcal{C}$. Define $\mathcal{F} = \bigcap_{i\in I} \mathcal{F}_i$. We know from a previous result that the intersection of an arbitrary number of σ-algebras is a σ-algebra. Therefore, $\mathcal{F} = \bigcap_{i\in I} \mathcal{F}_i$ is a σ-algebra.

Since each $\mathcal{F}_i$ contains $\mathcal{C}$ by definition, their intersection, $\mathcal{F}$, also contains $\mathcal{C}$. Thus, $\mathcal{C} \subseteq \mathcal{F}$.

Let $\sigma(\mathcal{C})$ denote the smallest σ-algebra containing $\mathcal{C}$. By definition, $\sigma(\mathcal{C})$ is a σ-algebra and contains $\mathcal{C}$. Therefore, it must be one of the $\mathcal{F}_i$ in the collection $\{\mathcal{F}_i\}_{i\in I}$. Since $\mathcal{F}$ is the intersection of all σ-algebras containing $\mathcal{C}$, it must be contained within any other σ-algebra that contains $\mathcal{C}$. In particular, $\mathcal{F} \subseteq \sigma(\mathcal{C})$.

Since $\mathcal{F}$ is defined as the intersection of all σ-algebras that contain $\mathcal{C}$, and $\sigma(\mathcal{C})$ itself is one of these σ-algebras, it follows that $\sigma(\mathcal{C}) \subseteq \mathcal{F}$. We have $\mathcal{F} \subseteq \sigma(\mathcal{C})$ and $\sigma(\mathcal{C}) \subseteq \mathcal{F}$. Therefore, $\mathcal{F} = \sigma(\mathcal{C})$.

This proves that the smallest σ-algebra containing $\mathcal{C}$ is $\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i$. □

We can now define the *Borel σ-algebra*. For this, we will have a setup - reasons of picking this setup will get clear in some time - when you will see that how open sets can be helpful in proving that even singleton sets are Borel sets.

Setup: Let $(0, 1]$ be the sample space, Ω. The collection of interesting sets of Ω, represented by $\mathcal{C}_0$, contains all open-intervals (a, b) in Ω.

Definition 2.1. *$\sigma(\mathcal{C}_0)$ is called the Borel σ-algebra, denoted by $\mathcal{B}((0, 1])$. An element of $\mathcal{B}((0, 1])$ is called Borel measurable set, or simply a Borel set.*

Thus, every open interval in $(0, 1]$ is a Borel set. We next prove that every singleton set in $(0, 1]$ is a Borel set too.

Theorem 2.4. *Every singleton set in $(0, 1]$ is a Borel set.*

Proof. Let $((0, 1], \mathcal{B})$ be the Borel space where $\mathcal{B}$ is the Borel σ-algebra generated by the open sets within $(0, 1]$. We want to show that any singleton set $\{x\}$, where $x \in (0, 1]$, is a Borel set.

Consider the singleton set $\{x\}$, where $x \in (0, 1]$. We can write it as:

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x + \frac{1}{n} \right) \cap (0, 1].$$

The above result can be proved by the method of contradiction. Let h be an element in $\bigcap_{n=1}^{\infty} \left(b - \frac{1}{n}, b + \frac{1}{n} \right)$ other than b. For every such h, there exists a large enough n_0 such that $h \notin \left(b - \frac{1}{n_0}, b + \frac{1}{n_0} \right)$. This implies $h \notin \bigcap_{n=1}^{\infty} \left(b - \frac{1}{n}, b + \frac{1}{n} \right)$.

Each interval $\left(x - \frac{1}{n}, x + \frac{1}{n} \right) \cap (0, 1]$ is an open set in $(0, 1]$, and since Borel sets are generated by open sets, these intervals belong to the Borel σ-algebra $\mathcal{B}$.

The Borel σ-algebra $\mathcal{B}$ is closed under countable intersections. Therefore, the intersection of the countable collection of open sets $\left(x - \frac{1}{n}, x + \frac{1}{n} \right) \cap (0, 1]$, which is exactly $\{x\}$, is also in $\mathcal{B}$.

Since $\{x\}$ can be expressed as a countable intersection of open sets within $(0, 1]$, it is a Borel set.

□

Corollary 2.1. *As an immediate consequence of this theorem, we see that every half-open interval, $(a, b]$, is a Borel set. This follows from the fact that*

$$(a, b] = (a, b) \cup \{b\},$$

and the fact that a countable union of Borel sets is a Borel set. For the same reason, every closed interval, $[a, b]$, is a Borel set.

This also gives one the idea of how to prove a set is a Borel set or not. If the set can be represented as a complement of an open set or as countable unions and countable intersections of open sets, it is a Borel set.

How big is the Borel σ-algebra?

Theorem 2.5. *The cardinality of the Borel σ-algebra (on the unit interval $(0,1]$) is the same as the cardinality of the reals. Thus, the Borel σ-algebra is a much smaller collection than the power set $2^{(0,1]}$.*

Proof. Let $\mathcal{B}$ denote the Borel σ-algebra on the unit interval $(0,1]$.

Step 1: Show that the cardinality of $\mathcal{B}$ is at most the cardinality of the reals.

The Borel σ-algebra $\mathcal{B}$ is generated by the open intervals of $(0,1]$, which form a basis for the topology. The set of all open intervals in $(0,1]$ has the same cardinality as $\mathbb{Q} \times \mathbb{Q}$, where $\mathbb{Q}$ is the set of rational numbers. Since $\mathbb{Q}$ is countable, it follows that the set of all open intervals in $(0,1]$ is also countable.

The Borel σ-algebra is generated by applying countable unions, intersections, and complements to these open intervals. Therefore, the number of sets that can be formed is bounded by $|\mathbb{R}|$, the cardinality of the reals.

Step 2: Show that the cardinality of $\mathcal{B}$ is at least the cardinality of the reals.

Consider the singleton sets $\{x\}$ where $x \in (0,1]$. Each singleton set is a Borel set, and the cardinality of these singleton sets is the same as the cardinality of the reals. Therefore, $\mathcal{B}$ must contain at least as many elements as the cardinality of $\mathbb{R}$.

Conclusion: We have shown that the cardinality of $\mathcal{B}$ is both at most and at least the cardinality of the reals. Therefore, the cardinality of the Borel σ-algebra $\mathcal{B}$ is exactly $|\mathbb{R}|$.

Since the power set $2^{(0,1]}$ has a cardinality of $2^{|\mathbb{R}|}$, which is strictly greater than $|\mathbb{R}|$, it follows that the Borel σ-algebra is a much smaller collection than the power set of $(0,1]$. □

2.1.3 What are not Borel sets?

The majority of sets in $(0,1]$ are Borel sets. In fact, the Borel σ-algebra on $(0,1]$ contains a wide range of sets, from simple open intervals to much more complex constructions. Identifying a non-Borel set is not trivial because the Borel σ-algebra is quite extensive.The Borel σ-algebra on $(0,1]$ includes many intricate sets such as the **Cantor set**. To understand the breadth of the Borel σ-algebra, we first prove that the Cantor set is a Borel set.

Lemma 2.1. *The Cantor Set is a Borel set.*

Proof. The Cantor set, C, is constructed by iteratively removing the middle third from each interval of $(0, 1]$.

$$C = \bigcap_{n=1}^{\infty} C_n,$$

where C_n is the set obtained after the n-th stage of removing the middle third of each interval.

Each C_n is a finite union of closed intervals. Since finite unions of closed sets are closed, C_n is closed for each n. The Cantor set C is the countable intersection of these closed sets. The Borel σ-algebra is closed under countable intersections of closed sets, so $C \in \mathcal{B}$, making the Cantor set a Borel set.

□

Examples of Non-Borel Sets

Although most familiar sets are Borel sets, there exist sets that are not Borel. These sets are usually constructed using the Axiom of Choice and involve more intricate arguments. One such example is the **Vitali set**. The Vitali set is constructed in the following way:

1. Consider the interval $[0, 1]$.
2. Define an equivalence relation $\sim$ on $[0, 1]$ by $x \sim y$ if and only if $x - y \in \mathbb{Q}$, i.e., x and y differ by a rational number.
3. By the Axiom of Choice, we select exactly one representative from each equivalence class under this relation. The collection of these representatives forms a set V, called the **Vitali set**.

But what is the Vitali Set actually?

Imagine you're organizing the quirkiest party ever on a number line between 0 and 1. This isn't just any party - it's a Vitali set party! Here's how you create your guest list:

First, you declare that two numbers are *dance partners* if their difference is a rational number. For example, 0.3 and 0.7 are dance partners because 0.7 - 0.3 = 0.4, which is rational. Now, you start grouping all the numbers between 0 and 1 into *dance troupes.* Each troupe consists of all numbers that are dance partners with each other. Here's the twist: you decide to invite exactly one person from each dance troupe to your party. It doesn't matter who you choose from each troupe, as long as you pick one and only one. The resulting guest list is what mathematicians call a Vitali set! Why is this party so special? Well, it has some mind-bending properties:

No two party guests are exactly one rational number apart. If Alice is at 0.3 and Bob is at 0.7, one of them didn't make the cut because they're in the same dance troupe. Yet, if you shift all your guests by any rational number, you'll get a completely new set of partiers, with no overlap with the original group! Despite seeming quite sparse (remember, we only chose one member from each dance troupe), this set has some very strange measuring properties, as we'll soon see.

Now that we've got our Vitali set party set up, let's explore why it's such a mathematical troublemaker!

Why is the Vitali Set Not a Borel Set?

Now, let's play a game called *Cover the Dance Floor.*

Here's how it works: We start with our Vitali set party on the (0,1] dance floor. We're given a magical dance move: we can shift everyone simultaneously by any rational number between -1 and 1. Our goal? Use these dance moves to cover every spot on a new, bigger dance floor from 0 to 2!

Here's the kicker: with the right series of these rational shifts, we can indeed cover every single point between 0 and 2. It's like our original Vitali set party has suddenly expanded to fill twice the space! But wait a minute... if the Vitali set were a Borel set (think of Borel sets as *well-behaved* sets that play nicely with measure theory), we'd run into a big problem. Here's why:

Borel sets have a property: if you take a Borel set and shift it by a rational number, the result is still a Borel set. We just covered the interval (0,2] using countably many rational shifts of our Vitali set. If the Vitali set were Borel, this new covered area would also be Borel. But here's the contradiction: we know the measure (think *length*) of (0,2] is 2, but it's made up of countably many copies of our original set, each of which should have the same measure as the original Vitali set.

Let's call the measure of the Vitali set m. Then we have:

$$2 = \text{countably many} \times m$$

This equation can't possibly work! If $m = 0$, the right side is zero. If $m > 0$, the right side is infinite. There's no value of m that makes this equation true. So, we're forced to conclude that our initial assumption - that the Vitali set is a Borel set - must be wrong. The Vitali set is too *wild* to be captured by the well-behaved Borel sets.

2.2 Introduction to Lebesgue Measure

What we understood from the example of *Vitali Set* in the last section is that we cannot assign it measure like *length*. We will only assign measures to the *Borel sets*. This gives us the understanding that the entire collection of subsets 2^{Ω}, where $\Omega = (0, 1]$, is not measureable. So, we can say $(\Omega, \mathcal{B})$ is the *measurable space*. Now, we want to assign each Borel set a measure.

Consider $\Omega = (0, 1]$. Let $\mathcal{F}_0$ consist of the empty set and all sets that are finite unions of intervals of the form $(a, b]$. A typical element of this set is of the form

$$F = (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n]$$

where $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_n < b_n$ and $n \in \mathbb{N}$.

Lemma 2.2. *(a) $\mathcal{F}_0$ is an algebra.*

(b) $\mathcal{F}_0$ is not a σ-algebra.

(c) $\sigma(\mathcal{F}_0) = \mathcal{B}$, where $\mathcal{B}$ denotes the Borel σ-algebra.

Proof. (a) By definition, $\emptyset \in \mathcal{F}_0$. Also, $\emptyset^C = (0, 1] \in \mathcal{F}_0$. The complement of $(a_1, b_1] \cup (a_2, b_2]$ is $(0, a_1] \cup (b_1, a_2] \cup (b_2, 1]$, which also belongs to $\mathcal{F}_0$. Furthermore, the union of finitely many sets, each of which is a finite union of intervals of the form $(a, b]$, is also a set that is a union of a finite number of intervals, and thus belongs to $\mathcal{F}_0$.

(b) To see this, note that $\left(0, \frac{n}{n+1}\right] \in \mathcal{F}_0$ for every n, but $\bigcup_{n=1}^{\infty} \left(0, \frac{n}{n+1}\right] = (0, 1) \notin \mathcal{F}_0$. This shows that $\mathcal{F}_0$ is not closed under countable unions, and thus it is not a σ-algebra.

(c) First, the null set is clearly a Borel set. Next, we have already seen that every interval of the form $(a, b]$ is a Borel set. Hence, every element of $\mathcal{F}_0$ (other than the null set), which is a finite union of such intervals, is also a Borel set. Therefore, $\mathcal{F}_0 \subseteq \mathcal{B}$. This implies $\sigma(\mathcal{F}_0) \subseteq \mathcal{B}$.

Next, we show that $\mathcal{B} \subseteq \sigma(\mathcal{F}_0)$. For any interval of the form (a, b) in $\mathcal{C}_0$, we can write

$$(a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right) \cap \Omega.$$

Since every interval of the form $\left(a, b - \frac{1}{n}\right) \in \mathcal{F}_0$, a countable union of such intervals belongs to $\sigma(\mathcal{F}_0)$. Therefore, $(a, b) \in \sigma(\mathcal{F}_0)$ and consequently, $\mathcal{C}_0 \subseteq \sigma(\mathcal{F}_0)$. This gives $\sigma(\mathcal{C}_0) \subseteq \sigma(\mathcal{F}_0)$. Using the fact that $\sigma(\mathcal{C}_0) = \mathcal{B}$ proves the required result.

□

Now, recall that we wanted to give subset (a, b) a measure that is proportional to $b - a$. While this makes sense for the intervals, it doesn't make sense for singleton sets and complex sets like *Cantor set*. What we want to do now is - extend the idea of this measure to other Borel sets. This is achieved by using **Caratheodory's Extension Theorem**.

Theorem 2.6. *Let $\mathcal{F}_0$ be an algebra of subsets of Ω, and let $\mathcal{F} = \sigma(\mathcal{F}_0)$ be the σ-algebra that it generates. Suppose that P_0 is a mapping from $\mathcal{F}_0$ to $[0, 1]$ that satisfies:*

1. *$P_0(\Omega) = 1$*
2. *P_0 is countably additive on $\mathcal{F}_0$.*

Then, P_0 can be extended uniquely to a probability measure on $(\Omega, \mathcal{F})$. That is, there exists a unique probability measure P on $(\Omega, \mathcal{F})$ such that $P(A) = P_0(A)$ for all $A \in \mathcal{F}_0$.

Proof. We proceed in several steps to establish the existence and uniqueness of the extension.

Step 1: Construction of an Outer Measure

Define an outer measure μ on the power set $\mathcal{P}(\Omega)$ as follows:

$$\mu(A) = \inf \left\{ \sum_{n=1}^{\infty} P_0(E_n) : A \subseteq \bigcup_{n=1}^{\infty} E_n,\ E_n \in \mathcal{F}_0 \right\}.$$

This definition uses the idea of covering A with a countable union of sets from $\mathcal{F}_0$. The sum of the measures of these covering sets provides an upper bound for $\mu(A)$.

The infimum ensures that we take the smallest possible value, making μ as small as possible while still covering A.

To show that μ is indeed an outer measure, we verify three properties:

- $\mu(\emptyset) = 0$ by definition.
- μ is monotone: if $A \subseteq B$, then $\mu(A) \leq \mu(B)$.
- μ satisfies countable subadditivity: for any sequence of sets $\{A_n\}_{n=1}^{\infty}$, $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

Step 2: Countable Additivity and Carathéodory's Extension Theorem

Since P_0 is countably additive on $\mathcal{F}_0$, it follows that μ is countably additive on $\mathcal{F}_0$. Specifically, for any sequence of disjoint sets $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}_0$, we have:

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} P_0(A_n).$$

This equality holds because the definition of μ coincides with P_0 on $\mathcal{F}_0$.

By Carathéodory's Extension Theorem, if an outer measure μ is countably additive on a collection of sets (here, $\mathcal{F}_0$), then μ can be extended to a measure P on the σ-algebra generated by those sets. Thus, there exists a unique measure P on $\mathcal{F}$ such that $P(A) = P_0(A)$ for all $A \in \mathcal{F}_0$.

Step 3: Verification of the Extension on $\mathcal{F}$

Since $\mathcal{F}$ is the σ-algebra generated by $\mathcal{F}_0$, every set in $\mathcal{F}$ can be expressed through countable unions, intersections, and complements of sets in $\mathcal{F}_0$. The measure P extends P_0 while preserving countable additivity. Therefore, for any $A \in \mathcal{F}_0$, we have:

$$P(A) = \mu(A) = P_0(A).$$

Step 4: Uniqueness of the Extension

Suppose there exists another probability measure Q on $(\Omega, \mathcal{F})$ that agrees with P_0 on $\mathcal{F}_0$. Let $A \in \mathcal{F}$. We can approximate A using sets from $\mathcal{F}_0$. Given that both P and Q agree with P_0 on $\mathcal{F}_0$, for any such approximation, the measures P and Q must produce the same value. Therefore:

$$P(A) \leq Q(A) \quad \text{and} \quad Q(A) \leq P(A).$$

This implies $P(A) = Q(A)$. Since A was arbitrary in $\mathcal{F}$, P and Q must be the same measure on $\mathcal{F}$.

Thus, P is the unique probability measure extending P_0 to $\mathcal{F}$.

$\square$

For every $F \in F_0$ of the form

$$F = (a_1, b_1] \cup (a_2, b_2] \cup \cdots \cup (a_n, b_n],$$

we define a function $P_0 : F_0 \to [0, 1]$ such that $P_0(\emptyset) = 0$ and

$$P_0(F) = \sum_{i=1}^{n} (b_i - a_i).$$

Note that $P_0(\Omega) = P_0((0, 1]) = 1$. Also, if $(a_1, b_1], (a_2, b_2], \ldots, (a_n, b_n]$ are disjoint sets, then

$$P_0\left(\bigcup_{i=1}^{n} (a_i, b_i]\right) = \sum_{i=1}^{n} P_0((a_i, b_i]) = \sum_{i=1}^{n} (b_i - a_i),$$

implying the finite additivity of P_0. It turns out that P_0 is countably additive on F_0 as well; that is, if $(a_1, b_1], (a_2, b_2], \ldots$ are disjoint sets such that $\bigcup_{i=1}^{\infty} (a_i, b_i] \in F_0$, then

$$P_0\left(\bigcup_{i=1}^{\infty} (a_i, b_i]\right) = \sum_{i=1}^{\infty} P_0((a_i, b_i]) = \sum_{i=1}^{\infty} (b_i - a_i).$$

From Carathéodory's extension theorem, there exists a unique probability measure P on $((0, 1], \mathcal{B})$ which is the same as P_0 on F_0. This unique probability measure on $(0, 1]$ is called the *Lebesgue* or *uniform measure*.

The Lebesgue measure formalizes the notion of length. Specifically, it extends the intuitive idea of length of intervals to a broader set of subsets of $\mathbb{R}$, including sets that are not necessarily intervals. The Lebesgue measure assigns to each set a non-negative value that represents its *size* in terms of length.

This suggests that the Lebesgue measure of a singleton should be zero. To demonstrate this, let $b \in (0, 1]$. Using the definition of the measure, we write

$$P(\{b\}) = P\left(\bigcap_{n=1}^{\infty} \left(b - \frac{1}{n}, b\right] \cap \Omega\right).$$

Let $A_n = \left(b - \frac{1}{n}, b\right]$. For each n, the Lebesgue measure of A_n is

$$P(A_n) = \frac{1}{n}.$$

Since $\{A_n\}$ is a decreasing sequence of nested sets,

$$P(\{b\}) = P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \frac{1}{n} = 0,$$

where the second equality follows from the continuity of probability measures.

Since any countable set is a countable union of singletons, the probability of a countable set is zero.

For example, under the uniform measure on $(0,1]$, the probability of the set of rationals is zero, since the rational numbers in $(0,1]$ form a countable set.

For $\Omega = (0,1]$, the Lebesgue measure is also a probability measure because $P((0,1]) = 1$. However, for other intervals (for example, $\Omega = (0,2]$), the Lebesgue measure is only a finite measure. In such cases, the measure can be normalized as appropriate to obtain a uniform probability measure. For instance, if $\Omega = (0,2]$, the Lebesgue measure of this interval is 2. By dividing by 2, we can create a uniform probability measure over $(0,2]$.

2.3 The Infinite Coin Toss Model

In this discussion, we explore a random experiment in which each trial involves infinitely many coin tosses. To make things simpler, let's represent each result of *Heads* and *Tails* with 0 and 1, respectively. In this setup, each sequence of outcomes from infinitely many tosses is represented by an infinite binary string. Thus, the sample space for this experiment can be described as

$$\Omega = \{0,1\}^\infty$$

where each outcome is a sequence of 0s and 1s extending infinitely. From *Real Analysis*, we know that such a space of all infinite binary sequences is uncountable. Therefore, defining a meaningful σ-algebra on Ω to handle probability requires careful consideration.

Let us introduce $\mathcal{F}_n$ as the collection of subsets of Ω that we can determine by observing the first n coin tosses alone. Formally, a subset $A \subset \Omega$ belongs to $\mathcal{F}_n$ if and only if there exists a subset $A^{(n)} \subset \{0,1\}^n$ such that:

$$A = \{\omega \in \Omega \mid (\omega_1, \omega_2, \ldots, \omega_n) \in A^{(n)}\}.$$

This means that membership in A depends solely on the first n values of any sequence in Ω.

Examples:

1. Let A_1 be the subset of Ω containing all sequences that have exactly two Heads in the first four tosses. Since A_1 depends only on the outcomes of the first four tosses, $A_1 \in \mathcal{F}_4$.
2. Let A_2 be the subset of Ω consisting of sequences where the third toss is a Head. Here, $A_2 \in \mathcal{F}_3$, since only the outcome of the first three tosses is needed to determine membership in A_2.

Observe that:

$$\mathcal{F}_n \subset \mathcal{F}_{n+1} \quad \forall n \in \mathbb{N}.$$

This inclusion indicates that as we increase the number of observed tosses, we can describe more subsets of Ω.

Although each $\mathcal{F}_n$ is indeed a σ-algebra, there is a limitation: it only allows us to describe subsets of Ω that can be resolved by observing a finite number of tosses. For instance, the set containing only the outcome where every toss results in Heads (an infinite sequence of 0s) is not in $\mathcal{F}_n$ for any finite n.

To address this limitation, we define the σ-algebra $\mathcal{F}_0$ as follows:

$$\mathcal{F}_0 = \bigcup_{i \in \mathbb{N}} \mathcal{F}_i.$$

In words, $\mathcal{F}_0$ represents the collection of all subsets of Ω that can be determined based on a finite number of coin tosses. Any subset in $\mathcal{F}_0$ must belong to $\mathcal{F}_i$ for some finite $i \in \mathbb{N}$.

Lemma 2.3. *We claim the following:*

1. $\mathcal{F}_0$ is an algebra.
2. $\mathcal{F}_0$ is not a σ-algebra.

Proof. **(i): $\mathcal{F}_0$ is an algebra**

To show that $\mathcal{F}_0$ is an algebra, we need to verify that it satisfies the following properties:

1. **Closure under finite unions:** If $A, B \in \mathcal{F}_0$, then $A \cup B \in \mathcal{F}_0$.
2. **Closure under finite intersections:** If $A, B \in \mathcal{F}_0$, then $A \cap B \in \mathcal{F}_0$.
3. **Closure under complements:** If $A \in \mathcal{F}_0$, then $A^c \in \mathcal{F}_0$.

Since $\mathcal{F}_0 = \bigcup_{i \in \mathbb{N}} \mathcal{F}_i$ and each $\mathcal{F}_i$ is a σ-algebra (and hence also an algebra), each individual $\mathcal{F}_i$ satisfies these closure properties. Because the elements of $\mathcal{F}_0$ are subsets of Ω whose membership can be decided by looking at only a finite number of coin tosses, we conclude that finite unions, intersections, and complements of such sets will also belong to $\mathcal{F}_0$. Therefore, $\mathcal{F}_0$ is an algebra.

(ii): $\mathcal{F}_0$ is not a σ-algebra

To show that $\mathcal{F}_0$ is not a σ-algebra, we need to find a countable collection of sets in $\mathcal{F}_0$ whose union or intersection does not belong to $\mathcal{F}_0$.

Consider the set

$$E = \{\omega \in \Omega \mid \text{every odd toss results in Heads}\}.$$

This set E is defined by an infinite condition, as it requires every odd-numbered toss in the sequence to result in a Head. Since the occurrence of E depends on an infinite number of tosses, it cannot be determined by observing any finite number of tosses.

Therefore, $E \notin \mathcal{F}_0$.

However, we can express E as a countable intersection of sets in $\mathcal{F}_0$ as follows:

$$E = \bigcap_{i=1}^{\infty} A_{2i-1},$$

where each $A_{2i-1} \in \mathcal{F}_0$ is the set of all binary strings with a Head at the $(2i-1)$-th position (i.e., each odd toss).

This example shows that $\mathcal{F}_0$ is not closed under countable intersections, which means $\mathcal{F}_0$ is not a σ-algebra.

To handle subsets like E, which require countable operations to be fully described, we define the smallest σ-algebra containing all the elements of $\mathcal{F}_0$, denoted by

$$\mathcal{F} = \sigma(\mathcal{F}_0).$$

This σ-algebra $\mathcal{F}$ includes all countable unions, intersections, and complements of sets in $\mathcal{F}_0$, thereby extending $\mathcal{F}_0$ to satisfy the properties of a σ-algebra.

□

2.3.1 A Probability Measure on $(\Omega = \{0,1\}^\infty, \mathcal{F})$

We will now define a probability measure on $\mathcal{F}$ that models the idea of a fair coin toss. The probability measure will be initially defined on a smaller collection $\mathcal{F}_0 \subset \mathcal{F}$, which contains events that are dependent only on a finite number of coin tosses. First, we define a finitely additive function P_0 on $\mathcal{F}_0$ such that $P_0(\Omega) = 1$. We will then extend P_0 to a full probability measure P on $\mathcal{F}$.

1. Defining P_0 on $\mathcal{F}_0$

For any event $A \in \mathcal{F}_0$, there exists some n such that $A \in \mathcal{F}_n$, where $\mathcal{F}_n$ is the collection of events determined by the outcomes of the first n coin tosses.

By the structure of $\mathcal{F}_n$, each event $A \in \mathcal{F}_n$ can be associated with a subset $A^{(n)} \subset \{0,1\}^n$, which represents the outcomes that define A after n tosses.

We define $P_0 : \mathcal{F}_0 \to [0,1]$ as:

$$P_0(A) = \frac{|A^{(n)}|}{2^n}$$

where $|A^{(n)}|$ is the number of outcomes in $A^{(n)}$.

2. Consistency of P_0 over n

We must ensure that the value of $P_0(A)$ is consistent, regardless of the choice of n in defining $A^{(n)}$. Since the collections $\mathcal{F}_n$ are nested (i.e., $\mathcal{F}_n \subset \mathcal{F}_{n+1}$), any event in $\mathcal{F}_n$ remains in $\mathcal{F}_{n+1}$, preserving the probability.

Example:

Consider the event A_2, which can be decided by the first three coin tosses. We have $A_2 \in \mathcal{F}_3$ and:

$$A^{(3)} = \{(0,0,0),(0,1,0),(1,0,0),(1,1,0)\} \Rightarrow |A^{(3)}| = 4$$

Then, $P_0(A_2) = \frac{4}{2^3} = \frac{1}{2}$.

When A_2 is considered in $\mathcal{F}_4$ (i.e., looking one toss further), $A^{(4)} = \{(0,0,0,0),(0,1,0,0),$ $(1,0,0,0),(1,1,0,0),(0,0,0,1),(0,1,0,1),(1,0,0,1),(1,1,0,1)\}$ with $|A^{(4)}| = 8$. Then:

$$P_0(A_2) = \frac{8}{2^4} = \frac{1}{2}$$

This example shows that P_0 remains consistent across different n, reinforcing the fairness of our coin-toss model.

3. Extending P_0 to a Probability Measure on $\mathcal{F}$

$P_0(\Omega) = 1$ and P_0 is finitely additive. Additionally, P_0 is countably additive on $\mathcal{F}_0$ (proof omitted here), allowing us to apply the Carathéodory extension theorem.

By this theorem, there exists a unique probability measure P on $(\Omega, \mathcal{F})$ that agrees with P_0 on $\mathcal{F}_0$.

4. Calculating the Probability of a New Event

Let E be the event where all odd-numbered tosses result in heads. Since $E \notin \mathcal{F}_0$, P_0 is not directly applicable. However, $E \in \mathcal{F}$, so P is defined for E.

Define $E_m = \bigcap_{i=1}^{m} A_{2i-1}$, where each $A_{2i-1} = \{\omega \in \Omega \mid \omega_{2i-1} = 0\}$.

We can compute:

$$P(E_m) = P_0(E_m) = \frac{1}{2^m}$$

since $\{E_m\}_{m\geq 1}$ forms a nested, decreasing sequence with $E = \bigcap_{m=1}^{\infty} E_m$.

Therefore:

$$P(E) = P\left(\bigcap_{m=1}^{\infty} E_m\right) = \lim_{m\to\infty} P(E_m) = \lim_{m\to\infty} \frac{1}{2^m} = 0$$

by the continuity of probability measures.

This completes the construction and verification of the probability measure P on $(\Omega, \mathcal{F})$ consistent with a fair coin-toss model.

Exercise 2.1. *Let us consider a collection of σ-algebras F_n for a fixed integer n. We notice that each F_n is limited, as it can only describe the outcomes of the first n coin tosses. We define a new σ-algebra F_0 as follows:*

$$F_0 = \bigcap_{i=1}^{\infty} F_n.$$

Provide a verbal description of the collection F_0.

Solution 2.1. *To understand the collection F_0, let's break it down:*

1. Each σ-algebra F_n captures all possible events that can occur from observing the first n tosses of a coin. This includes events such as:

The outcome of each individual toss (Heads or Tails).
The total number of Heads or Tails in those n tosses.
Any combination of these outcomes.

2. When we define F_0 as the union of all F_n for $n = 1, 2, 3, \ldots$, we are essentially saying that F_0 encompasses all possible events from any number of coin tosses, not just the first n.

3. Therefore, F_0 includes events such as:

The outcome of any finite number of tosses.
The total number of Heads or Tails from an infinite sequence of tosses.
Any event that could be defined by the outcomes of infinitely many tosses.

In summary, the collection F_0 represents the σ-algebra that contains all events that can be formed from an infinite sequence of coin tosses, allowing us to model every possible scenario that could arise from tossing the coin infinitely many times.

Exercise 2.2. *Show that $\mathcal{F}_0$ is an algebra on Ω.*

Solution 2.2. *To demonstrate that $\mathcal{F}_0$ is an algebra on Ω, we must verify that it satisfies three essential properties:*

*1. **Containment of the Sample Space**: We must show that the entire sample space Ω is included in $\mathcal{F}_0$. By definition, an algebra requires that the sample space be one of the elements within it.*

*2. **Closure under Complements**: If $A \in \mathcal{F}_0$, we need to establish that the complement of A, denoted by A^c, is also in $\mathcal{F}_0$. This is crucial because an algebra must contain not only its elements but also the elements that are not in those sets.*

*3. **Closure under Finite Unions**: For any two sets $A, B \in \mathcal{F}_0$, we need to show that their union $A \cup B$ also belongs to $\mathcal{F}_0$. The algebra property requires that the combination of sets remains within the structure of the algebra.*

Hence, $\mathcal{F}_0$ is an algebra on Ω.

Exercise 2.3. *Consider the subset $A \subset \Omega$ consisting of sequences in which Tails occurs infinitely many times. Does $A \in \mathcal{F}_0$? Is A^c countable?*

Solution 2.3. *To understand the problem, we first define the set A. This set consists of all sequences of coin tosses where Tails appears infinitely often. In simpler terms, if we flip a coin repeatedly, the sequences in A will have Tails show up no matter how far we go in the sequence.*

Next, we need to determine if A is a member of the σ-algebra $\mathcal{F}_0$. The σ-algebra is a collection of events we can measure and is generated by the basic outcomes of the experiment, such as the outcomes of individual coin tosses.

To explore this, consider the infinite nature of the sequences in A. For A to belong to $\mathcal{F}_0$, it must be possible to express it using the basic events available in $\mathcal{F}_0$, which typically involve finitely many tosses. However, the characteristic of A being defined by the occurrence of Tails infinitely often makes it a more complex event than can be simply constructed from finite tosses.

Now, let's analyze the complement of A, denoted as A^c. The set A^c consists of sequences where Tails occurs only a finite number of times. For example, a sequence in A^c might look like HHHHT (where Tails only occurs once), or it could be HHHHHTHH (where Tails occurs twice).

To see if A^c is countable, we note that each sequence in A^c can be described by the finite number of Tails and the positions in which they appear among an infinite number of Heads. Since there are only finitely many choices for where to place Tails in a sequence, we can represent each sequence in A^c with a finite binary string.

In conclusion, A cannot be measured within the simple framework of $\mathcal{F}_0$, while A^c is countable due to its finite nature.

Exercise 2.4. *Let B be the set of all infinite sequences for which $\omega_n = 0$ for every odd n; that is, every odd-numbered toss results in Heads. Show that B can be written as a countable intersection of subsets in $\mathcal{F}_0$, but $B \notin \mathcal{F}_0$. Therefore, $\mathcal{F}_0$ is not a σ-algebra. Define $\mathcal{F} = \sigma(\mathcal{F}_0)$, the σ-algebra generated by $\mathcal{F}_0$.*

The elements of B are infinite sequences $\omega = (\omega_1, \omega_2, \omega_3, \ldots)$, where each odd-indexed position ω_n equals zero. In simpler terms, every first, third, fifth, etc., toss results in Heads.

Now, we need to show that B can be represented as a countable intersection of subsets within $\mathcal{F}_0$.

Solution 2.4. *Consider the sets A_k defined as follows:*

$$A_k = \{\omega \in \mathcal{F}_0 : \omega_{2k-1} = 0\}$$

for $k = 1, 2, 3, \ldots$. Each A_k represents the set of sequences where the $(2k-1)$-th toss is Heads.

Thus, we can express the set B as:

$$B = \bigcap_{k=1}^{\infty} A_k$$

This is because B requires that all odd-indexed tosses (which correspond to $\omega_1, \omega_3, \omega_5, \ldots$) are equal to zero. Each set A_k is in $\mathcal{F}_0$, and since B is a countable intersection of sets in $\mathcal{F}_0$, it follows that B can be expressed as such.

Next, we need to demonstrate that $B \notin \mathcal{F}_0$. If B were to belong to $\mathcal{F}_0$, it would imply that $\mathcal{F}_0$ is closed under countable intersections. However, this contradicts the definition of a σ-algebra, which must contain all countable intersections of its sets.

Thus, we conclude that $\mathcal{F}_0$ cannot be a σ-algebra since it does not contain the intersection B.

Lastly, we define $\mathcal{F}$ as follows:

$$\mathcal{F} = \sigma(\mathcal{F}_0)$$

This σ-algebra $\mathcal{F}$ is the smallest σ-algebra containing $\mathcal{F}_0$.

In summary, while $\mathcal{F}_0$ does not qualify as a σ-algebra due to the exclusion of the intersection B, the generated σ-algebra $\mathcal{F}$ encompasses all necessary sets, including B.

Exercise 2.5. *Show that every singleton $\{\omega\}$ is $\mathcal{F}$-measurable. Show that the uniform measure on $(\Omega, \mathcal{F})$ defined in class assigns zero probability measure to singletons.*

Solution 2.5. *To demonstrate that every singleton $\{\omega\}$ is $\mathcal{F}$-measurable, we need to establish that for any singleton $\{\omega\}$, the set is included in the σ-algebra $\mathcal{F}$. By the definition of a σ-algebra, it contains all sets formed by countable unions, intersections, and complements of the sets in $\mathcal{F}_0$.*

Since $\mathcal{F}$ is generated from $\mathcal{F}_0$, and $\mathcal{F}_0$ contains sets based on outcomes from our probability space (including sequences of coin tosses), we can assert that:

$$\{\omega\} \in \mathcal{F}$$

Thus, every singleton $\{\omega\}$ is $\mathcal{F}$-measurable.

Now, let's consider the uniform measure P defined on $(\Omega, \mathcal{F})$. The uniform measure assigns probabilities in a way that is equally distributed across all possible outcomes in the sample space Ω.

Since Ω is an infinite set, we can define the uniform measure as follows:

$$P(A) = \frac{\text{number of elements in } A}{\text{number of elements in } \Omega}$$

For any singleton $\{\omega\}$, we have:

$$P(\{\omega\}) = \frac{1}{\infty} = 0$$

This result indicates that the measure assigned to any singleton $\{\omega\}$ is zero. Consequently, we conclude that the uniform measure on $(\Omega, \mathcal{F})$ assigns zero probability to singletons.

Exercise 2.6. *Let A_i be the set of all outcomes such that the i-th toss is Tails. Note that $A_i \in \mathcal{F}_0$. Show that the set A can be written as*

$$A = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$$

Hence, show that A is $\mathcal{F}$-measurable. What is $P(A)$ under the uniform measure?

Solution 2.6. *Let A_i be the set of all outcomes such that the i-th toss is Tails. It is important to note that $A_i \in \mathcal{F}_0$.*

We aim to show that the set A can be expressed as follows:

$$A = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$$

The expression above conveys that the event A occurs if there exists at least one n such that all tosses from the n-th toss onward are Tails.

1. The inner intersection $\bigcap_{i=n}^{\infty} A_i$ captures the outcome where all tosses starting from the n-th toss are Tails.
2. The outer union $\bigcup_{n=1}^{\infty}$ indicates that we are considering this event for every possible starting point n.

Since A_i is in $\mathcal{F}_0$ for each i, and since $\mathcal{F}_0$ is closed under countable unions and intersections, it follows that A is a combination of the sets A_i using these operations.

Therefore, A is measurable with respect to $\mathcal{F}$.

Under the uniform measure, we can determine $P(A)$ as follows:

Since each toss of a fair coin results in Tails with probability $\frac{1}{2}$, the probability that an infinite sequence of tosses results in Tails from the n-th toss onward is given by:

$$P\left(\bigcap_{i=n}^{\infty} A_i\right) = \left(\frac{1}{2}\right)^{\infty} = 0$$

Thus, the probability of A can be calculated by considering the probability that there exists some n such that all tosses from n onward are Tails.

However, because we are considering the complement (the event where not all outcomes from n onward are Tails), we have:

$$P(A) = 1 - P\left(\bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i\right) = 1 - 0 = 1$$

Hence, we conclude that:

$$P(A) = 1$$

Exercise 2.7. *Let $T \subseteq \Omega$ be the set of all coin toss sequences in which the fraction of Tails is exactly $\frac{1}{2}$. More precisely, we define T as follows:*

$$T = \{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \omega_i = \frac{1}{2}\}$$

The set T is called the strong-law truth set, for reasons that will become clear later. Does $T \in \mathcal{F}_0$?

Solution 2.7. *To determine whether $T \in \mathcal{F}_0$, we need to analyze the nature of this set. The condition for membership in T involves taking the limit of a ratio as n approaches infinity, which is a property based on the convergence of the sequence ω.*

In general, $\mathcal{F}_0$ contains sets that can be described by finite combinations of events (i.e., measurable sets) but may not include all possible limit points or convergence properties defined in terms of sequences.

Since T requires the evaluation of an infinite limit and depends on the behavior of the entire sequence, it is not typically captured by the events in $\mathcal{F}_0$, which often encompass finite or countably infinite unions and intersections of more elementary sets.

Thus, we conclude that:

$$T \notin \mathcal{F}_0$$

Exercise 2.8. *Show that T can be expressed as:*

$$T = \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \left\{ \omega \in \Omega \middle| \left| \frac{1}{n} \sum_{i=1}^{n} \omega_i - \frac{1}{2} \right| < \frac{1}{k} \right\}$$

Argue that the subset inside the nested union and intersection above belongs to $\mathcal{F}_0$.

Solution 2.8. *To do so, we need to examine the structure of the sets involved. The set*

$$\left\{ \omega \in \Omega \middle| \left| \frac{1}{n} \sum_{i=1}^{n} \omega_i - \frac{1}{2} \right| < \frac{1}{k} \right\}$$

represents sequences of outcomes for which the average of the first n outcomes converges to $\frac{1}{2}$ as n becomes large. This is a condition that can be expressed in terms of finite sequences and their sums, which are measurable with respect to $\mathcal{F}_0$.

Now, let's show that T is $\mathcal{F}$-measurable. We start by rewriting the definition of T: The set T can be expressed as the set of all $\omega \in \Omega$ such that for all $k \geq 1$, there exists an m such that for all $n > m$:

$$\left| \frac{1}{n} \sum_{i=1}^{n} \omega_i - \frac{1}{2} \right| < \frac{1}{k}$$

This means that T consists of those sequences for which the average of the outcomes converges to $\frac{1}{2}$ as n goes to infinity.

Since the conditions defining the sets in T involve countable unions and intersections of measurable sets from $\mathcal{F}_0$, and since $\mathcal{F}$ is closed under these operations, we conclude that T is indeed $\mathcal{F}$-measurable.

Therefore, we have shown that T can be expressed in the stated form and is measurable with respect to the σ-algebra $\mathcal{F}$.

Chapter 3

Conditional Probability and Independence

After the rigorous treatment of the *probability measure* in the last chapter, we can pick up a lot of the results of general probability from the *Classical Probability*. So far we have considered - *what probability spaces are and how probability measures are assigned*. In this chapter, we will have a look at the *Conditional Probability*.

3.1 Introduction to Conditional Probability

Definition 3.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is the sample space, $\mathcal{F}$ is a σ-algebra, and P is a probability measure. If we have an event B that belongs to the σ-algebra $\mathcal{F}$ and satisfies $P(B) > 0$, we can define the conditional probability of an event A given B using the following formula:*

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

This expression tells us how likely event A is to occur when we know that event B has occurred.

It is important to note that we cannot condition our probabilities on sets that have a zero probability measure. For instance, consider the probability space where $\Omega = [0, 1]$ with the Borel σ-algebra and a uniform probability measure. In this context, the set of rational numbers, which is countable, has a probability measure of zero. Therefore, it would be inappropriate to condition on this set of rational numbers when calculating probabilities.

Theorem 3.1. *Let $B \in \mathcal{F}$ and $P(B) > 0$. Then, the function $P(\cdot|B) : \mathcal{F} \to [0, 1]$ is a probability measure on the measurable space $(\Omega, \mathcal{F})$. $(\cdot)$ means you can take argument from $\mathcal{F}$.*

Proof. To establish that $P(\cdot|B)$ is indeed a probability measure, we must verify three key properties of probability measures:

1. $P(\Omega|B) = 1$.
2. $P(\varnothing|B) = 0$.
3. The property of countable additivity.

We begin by proving the first property:

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Next, we prove the second property:

$$P(\varnothing|B) = \frac{P(\varnothing \cap B)}{P(B)} = \frac{P(\varnothing)}{P(B)} = 0.$$

Now, we focus on proving the countable additivity property. Let $A_1, A_2, \ldots$ be a sequence of disjoint events. We need to demonstrate that

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

Consider the left-hand side:

$$P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) \cap B\right)}{P(B)}.$$

Using the properties of unions and the fact that the events A_i are disjoint, we can express this as:

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

Since the events A_i are disjoint, it follows that $A_i \cap B$ are also disjoint. Hence, we can apply the property of probability measures to the disjoint union:

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

Thus, we have:

$$P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i|B).$$

This completes the proof that $P(\cdot|B)$ satisfies the countable additivity property. Therefore, we conclude that $P(\cdot|B)$ is indeed a probability measure on $(\Omega, \mathcal{F})$. □

3.1.1 Properties of Conditional Probability

Theorem 3.2. ***The Law of Total Probability.*** *The Law of Total Probability states that if we have an event A from a σ-algebra $\mathcal{F}$ and a collection of events $\{B_i\}_{i=1}^{\infty}$ that form a partition of the sample space Ω, this means two things:*

1. The union of all events B_i covers the entire sample space, i.e.,

$$\bigcup_{i \in \mathbb{N}} B_i = \Omega.$$

2. No two events B_i and B_j can occur at the same time if $i \neq j$, which is expressed mathematically as:

$$B_i \cap B_j = \emptyset \quad \text{for all } i \neq j.$$

We also require that the probability of each B_i is greater than zero, i.e., $P(B_i) > 0$ for all i.

Given these conditions, the probability of the event A can be computed using the formula:

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i).$$

This equation tells us that to find the total probability of A, we sum the probabilities of A occurring given each B_i, weighted by the probability of each B_i.

Proof. Since the events $\{B_i\}$ partition Ω, the intersection of A with each B_i also partitions A. Therefore, by the property of countable additivity, we have:

$$P(A) = P\left(\bigcup_{i=1}^{\infty}(A \cap B_i)\right) = \sum_{i=1}^{\infty} P(A \cap B_i).$$

Using the definition of conditional probability, we know:

$$P(A \cap B_i) = P(A|B_i)P(B_i) \quad \text{for all } i.$$

Substituting this into our equation, we get:

$$\sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i).$$

As a special case, if we have an event B such that $0 < P(B) < 1$, we can express $P(A)$ as:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c),$$

where B^c is the complement of B, indicating the events that are not part of B. □

Theorem 3.3. ***Bayes' Rule.*** *Consider an event A in a σ-algebra $\mathcal{F}$ with $P(A) > 0$ and a collection of events $\{B_i\}_{i=1}^{\infty}$ forming a partition of the sample space Ω where $P(B_i) > 0$ for all i. Bayes' Rule states that*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{\infty} P(A|B_j)P(B_j)}.$$

Proof. To prove this, we start from the definition of conditional probability:

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

Next, we express $P(A)$ as:

$$P(A) = \sum_{j=1}^{\infty} P(A \cap B_j) = \sum_{j=1}^{\infty} P(B_j)P(A|B_j).$$

Substituting this into our equation gives:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{\infty} P(A|B_j)P(B_j)}.$$

Thus, Bayes' Rule allows us to update the probability of B_i given that A has occurred. □

Theorem 3.4. *For any sequence of events $\{A_i\}$, we can express the probability of the occurrence of at least one of these events as follows:*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(A_1) \cdot \prod_{i=2}^{\infty} P(A_i|A_1 \cap A_2 \cap \ldots \cap A_{i-1}),$$

provided that all the conditional probabilities are well-defined.

Proof. To understand why this holds, let's first consider a finite set of events. For n events, we have:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_1) \cdot \prod_{i=2}^{n} P(A_i|A_1 \cap A_2 \cap \ldots \cap A_{i-1}).$$

Now, if we take the limit as n approaches infinity, we write:

$$\lim_{n\to\infty} P\left(\bigcup_{i=1}^{n} A_i\right) = \lim_{n\to\infty} P(A_1) \cdot \prod_{i=2}^{n} P(A_i|A_1 \cap A_2 \cap \ldots \cap A_{i-1}).$$

Using the continuity of probability, we find that we can interchange the limit and the product, leading us to the desired relationship:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(A_1) \cdot \prod_{i=2}^{\infty} P(A_i|A_1 \cap A_2 \cap \ldots \cap A_{i-1}).$$

This formulation allows us to break down the complex event of an infinite union into a product of probabilities, illustrating the interconnectedness of the events involved. □

3.2 Independence

Definition 3.2. *Consider a probability space denoted by $(\Omega, \mathcal{F}, P)$. We say that two events A and B are independent with respect to the probability measure P if the following condition holds:*

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

It is important to note that if $P(B) > 0$ and the events A and B are independent, we can derive:

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

Example: We should consider whether disjoint sets can ever be independent. Let A and B be two disjoint events in $\mathcal{F}$. By the definition of disjoint events, we have:

$$P(A \cap B) = P(\emptyset) = 0.$$

For A and B to be independent, we require:

$$P(A \cap B) = P(A)P(B) = 0.$$

This condition can only be satisfied if either $P(A) = 0$ or $P(B) = 0$. Therefore, we conclude that two disjoint events are independent if and only if at least one of them has a probability of zero.

Definition 3.3. *A collection of events $A_1, A_2, \ldots, A_n$ is said to be independent if, for any non-empty subset $I_0 \subseteq \{1, 2, \ldots, n\}$, the following relationship holds:*

$$P\left(\bigcap_{i \in I_0} A_i\right) = \prod_{i \in I_0} P(A_i).$$

Next, we extend this concept to an arbitrary collection of events.

Definition 3.4. *A collection of events $\{A_i, i \in I\}$ is defined to be independent if, for every non-empty finite subset I_0 of I, the equality holds:*

$$P\left(\bigcap_{i \in I_0} A_i\right) = \prod_{i \in I_0} P(A_i).$$

3.2.1 Independence of σ-algebra

Definition 3.5. *Let F_1 and F_2 be two sub-σ-algebras of F. We say that F_1 and F_2 are independent σ-algebras if for every event $A_1 \in F_1$ and $A_2 \in F_2$, the events A_1 and A_2 are independent.*

Example: A straightforward example can be constructed as follows: Let $A, B \in F$. Define $F_1 = \{\emptyset, \Omega, A, A^c\}$ and $F_2 = \{\emptyset, \Omega, B, B^c\}$. The σ-algebras F_1 and F_2 are independent if and only if the events A and B are independent.

Next, we extend the concept of independence to a collection of sub-σ-algebras.

Definition 3.6. *Let $\{F_i, i \in I\}$ be a collection of sub-σ-algebras of F, where I is an index set. The collection $\{F_i, i \in I\}$ is said to be independent if for any choice of events $A_i \in F_i$, the events $\{A_i, i \in I\}$ are independent.*

Example: Consider the infinite coin toss model we discussed previously.

Let A_i be the event that the i-th coin toss resulted in heads. If $i \neq j$, then the events A_i and A_j are independent. This implies that the infinite collection of events $\{A_i \mid i \in \mathbb{N}\}$ is independent, capturing the intuitive idea of independent coin tosses.

Now, let F_1 (respectively, F_2) denote the collection of all events whose occurrence can be determined by examining the results of coin tosses at odd times (respectively, at even times).

Formally, define H_i as the event that the i-th toss resulted in heads. Set $C = \{H_i \mid i \text{ is odd}\}$, and let $F_1 = \sigma(C)$, which is the smallest σ-algebra containing all the events H_i for odd i. Similarly, we define F_2 using the tosses at even times.

In this context, the two σ-algebras F_1 and F_2 are independent. This means that any event determined solely by the outcomes of tosses at odd times is independent of any event determined solely by tosses at even times.

Lastly, let F_n be the collection of all events that can be determined by examining the coin tosses $2n$ and $2n+1$. It is known that F_n is a σ-algebra with finitely many events for every $n \in \mathbb{N}$. Remarkably, the collections $\{F_n, n \in \mathbb{N}\}$ are also independent.

Exercise 3.1. *Let C and D be elements of a sigma-algebra $\mathcal{F}$ on a sample space Ω. We aim to show that the collections $\mathcal{F}_1 = \{\varnothing, \Omega, C, C^c\}$ and $\mathcal{F}_2 = \{\varnothing, \Omega, D, D^c\}$ are independent if and only if C and D are independent events.*

Solution 3.1. *The concept of independence of events in probability is extended to the independence of collections of sets. In this problem, we have:*

1. A sigma-algebra $\mathcal{F}$, which is a collection of sets that is closed under complementation and countable unions.
2. Two specific collections of sets, $\mathcal{F}_1$ and $\mathcal{F}_2$, generated by the events C and D and their complements.

Our task is to show that the independence of the collections $\mathcal{F}_1$ and $\mathcal{F}_2$ is equivalent to the independence of the events C and D.

Two collections of sets, $\mathcal{F}_1$ and $\mathcal{F}_2$, are said to be independent if for every $A \in \mathcal{F}_1$ and every $B \in \mathcal{F}_2$, the probability of their intersection satisfies:

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

We need to check whether all combinations of elements in $\mathcal{F}_1$ and $\mathcal{F}_2$ satisfy the independence condition. However, many of these combinations are trivial:

1. For example, $P(\varnothing \cap B) = 0 = P(\varnothing) \cdot P(B)$ for any $B \in \mathcal{F}_2$, and similarly for $A \cap \varnothing$.
2. Also, for $A = \Omega$ or $B = \Omega$, independence holds as $P(\Omega \cap B) = P(B)$ and $P(A \cap \Omega) = P(A)$.

The non-trivial cases are when $A = C$ or C^c and $B = D$ or D^c. We check each of these cases:

For $A = C$ and $B = D$:

$$P(C \cap D) = P(C) \cdot P(D) \quad \textit{(this is the definition of independence for events C and D)}$$

For $A = C$ and $B = D^c$:

$$P(C \cap D^c) = P(C) \cdot P(D^c).$$

Since $P(D^c) = 1 - P(D)$, this follows from the independence of C and D.

For $A = C^c$ and $B = D$:
$$P(C^c \cap D) = P(C^c) \cdot P(D),$$
and again, this follows from the independence of C and D, since $P(C^c) = 1 - P(C)$.

For $A = C^c$ and $B = D^c$:
$$P(C^c \cap D^c) = P(C^c) \cdot P(D^c).$$
This completes the verification for all possible non-trivial cases.

Thus, we have shown that the collections $\mathcal{F}_1 = \{\varnothing, \Omega, C, C^c\}$ and $\mathcal{F}_2 = \{\varnothing, \Omega, D, D^c\}$ are independent if and only if the events C and D are independent.

Exercise 3.2. *Let $\Omega = \{1, 2, 3, \ldots, p\}$, where p is a prime number. Let $\mathcal{F}$ be the collection of all subsets of Ω, and define a probability measure P on events $A \in \mathcal{F}$ by*

$$P(A) = \frac{|A|}{p},$$

where $|A|$ denotes the cardinality of A. Show that if A and B are independent events, then at least one of A and B is either $\varnothing$ or Ω.

Solution 3.2. *To prove this, we will explore the properties of independent events A and B under the probability measure P defined on subsets of Ω.*

By definition, two events A and B are independent if

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

According to our probability measure, this means

$$\frac{|A \cap B|}{p} = \frac{|A|}{p} \cdot \frac{|B|}{p}.$$

Simplifying this equation, we get

$$|A \cap B| = \frac{|A| \cdot |B|}{p}.$$

The equation above indicates that the cardinality $|A \cap B|$ is equal to $\frac{|A| \cdot |B|}{p}$. Since $|A|$, $|B|$, and $|A \cap B|$ are all integers, $\frac{|A| \cdot |B|}{p}$ must also be an integer.

Here, p is a prime number. Therefore, for $\frac{|A| \cdot |B|}{p}$ to be an integer, p must divide the product $|A| \cdot |B|$. This divisibility can only happen if one of the following is true:

$|A| = 0$ *(which implies $A = \varnothing$),*
$|B| = 0$ *(which implies $B = \varnothing$),*
$|A| = p$ *(which implies $A = \Omega$),*
$|B| = p$ *(which implies $B = \Omega$).*

Thus, for A and B to satisfy the independence condition under the given probability measure, at least one of A and B must be either the empty set $\varnothing$ or the entire set Ω.

Exercise 3.3. *In a box, there are four red balls, six red cubes, six blue balls, and an unknown number of blue cubes. When an object from the box is selected at random, the shape and colour of the object are independent. Determine the number of blue cubes.*

Solution 3.3. *Let us define the following quantities to represent the count of each object in the box:*

- *Let R_B be the number of red balls, so $R_B = 4$.*
- *Let R_C be the number of red cubes, so $R_C = 6$.*
- *Let B_B be the number of blue balls, so $B_B = 6$.*
- *Let B_C be the unknown number of blue cubes, which we need to determine.*

Let N be the total number of objects in the box:

$$N = R_B + R_C + B_B + B_C = 4 + 6 + 6 + B_C = 16 + B_C$$

*Now, we are told that the shape (ball or cube) and the colour (red or blue) of an object are **independent**. This means that the probability of selecting an object of a particular shape is independent of the probability of selecting an object of a particular colour.*

Let's compute the probabilities of selecting each shape and each colour independently:

- *The probability of selecting a ball (regardless of colour) is given by:*

$$P(ball) = \frac{R_B + B_B}{N} = \frac{4+6}{16+B_C} = \frac{10}{16+B_C}$$

- *The probability of selecting a cube (regardless of colour) is given by:*

$$P(cube) = \frac{R_C + B_C}{N} = \frac{6+B_C}{16+B_C}$$

- *The probability of selecting a red object (regardless of shape) is:*

$$P(red) = \frac{R_B + R_C}{N} = \frac{4+6}{16+B_C} = \frac{10}{16+B_C}$$

- *The probability of selecting a blue object (regardless of shape) is:*

$$P(blue) = \frac{B_B + B_C}{N} = \frac{6+B_C}{16+B_C}$$

For independence to hold, the probability of selecting a ball that is red, $P(ball \cap red)$, should equal the product of the probabilities of selecting a ball and selecting a red object:

$$P(ball \cap red) = P(ball) \cdot P(red)$$

Now, calculating $P(ball \cap red)$:

$$P(ball \cap red) = \frac{R_B}{N} = \frac{4}{16+B_C}$$

And calculating $P(ball) \cdot P(red)$:

$$P(ball) \cdot P(red) = \frac{10}{16 + B_C} \cdot \frac{10}{16 + B_C} = \frac{100}{(16 + B_C)^2}$$

Setting these two expressions equal for independence, we get:

$$\frac{4}{16 + B_C} = \frac{100}{(16 + B_C)^2}$$

Cross-multiplying to solve for B_C, we obtain:

$$4(16 + B_C) = 100$$

$$64 + 4B_C = 100$$

$$4B_C = 36$$

$$B_C = 9$$

Exercise 3.4. *A man is known to speak the truth $\frac{3}{4}$ of the time. He throws a die and reports that it is a six. Find the probability that it is actually a six.*

Solution 3.4. *Let's solve this problem using Bayes' theorem, which allows us to find the probability of an event given some conditional information about it.*

Let A be the event that the die shows a six, and B be the event that the man reports a six.

We want to find $P(A|B)$, the probability that the die actually shows a six given that the man reports it as a six.

Using Bayes' theorem, we know:

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

Since a fair die has six faces, the probability of rolling a six is:

$$P(A) = \frac{1}{6}$$

This is the probability that the man reports a six given that a six has actually occurred. Since the man tells the truth $\frac{3}{4}$ of the time:

$$P(B|A) = \frac{3}{4}$$

Here, $\neg A$ is the event that the die shows something other than a six. If the die does not show a six, the probability that the man reports a six (i.e., he lies) is $\frac{1}{4}$:

$$P(B|\neg A) = \frac{1}{4}$$

The probability that the die does not show a six is:

$$P(\neg A) = 1 - P(A) = 1 - \frac{1}{6} = \frac{5}{6}$$

We can now find $P(B)$ using the law of total probability:

$$P(B) = P(B|A) \cdot P(A) + P(B|\neg A) \cdot P(\neg A)$$

Substituting the values, we get:

$$P(B) = \frac{3}{4} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{5}{6}$$

$$= \frac{3}{24} + \frac{5}{24} = \frac{8}{24} = \frac{1}{3}$$

Substituting all of the values into Bayes' theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} = \frac{\frac{3}{4} \cdot \frac{1}{6}}{\frac{1}{3}}$$

$$= \frac{\frac{3}{24}}{\frac{1}{3}} = \frac{3}{24} \cdot 3 = \frac{3}{8}$$

Exercise 3.5. *Let A and B be two events in a probability space with probabilities $P(A|B) > P(A)$. This inequality implies that the probability of A occurring, given that B has already occurred, is greater than the probability of A occurring independently of B. We aim to show the following:*

1. $P(B|A) > P(B)$
2. $P(A|\text{complement of } B) < P(A)$

Solution 3.5. *By Bayes' theorem, we know that:*

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

and similarly,

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

Given that $P(A|B) > P(A)$, we can rewrite this as:

$$\frac{P(A \cap B)}{P(B)} > P(A).$$

Rearranging this inequality, we obtain:

$$P(A \cap B) > P(A) \cdot P(B).$$

Now, using Bayes' theorem again for $P(B|A)$, we substitute $P(A \cap B)$ from the inequality above:

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

Therefore, we have shown that $P(B|A) > P(B)$.

Since B^c represents the complement of B, we use the law of total probability. According to this law, we can express $P(A)$ as:

$$P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c).$$

Now, because we are given $P(A|B) > P(A)$, it follows that $P(A|B^c)$ must be less than $P(A)$ to balance the equation. If $P(A|B^c)$ were not less than $P(A)$, then $P(A)$ would not satisfy the inequality required by the law of total probability, given that $P(A|B)$ is already greater than $P(A)$.

Thus, we conclude that:

$$P(A|B^c) < P(A).$$

Exercise 3.6. *A coin is tossed independently n times, with the probability of heads in each toss being p. At each time k (for $k = 2, 3, \ldots, n$), we get a reward at time $k + 1$ if the k-th toss results in a head and the $(k-1)$-th toss was a tail. Let A_k be the event that a reward is obtained at time k.*

1. *Are the events A_k and A_{k+1} independent?*

2. *Are the events A_k and A_{k+2} independent?*

Solution 3.6. *To answer each part, let's first explore the nature of event A_k in terms of the toss sequence. Since A_k represents obtaining a reward at time k, this implies that:*

1. The toss at time $k - 1$ must be a tail.
2. The toss at time k must be a head.

*Thus, A_k occurs when the outcome sequence for tosses $(k - 1, k)$ is **Tail-Head**. This outcome has a probability of $(1 - p)p$.*

For events A_k and A_{k+1} to be independent, the occurrence of A_k should not affect the probability of A_{k+1} occurring.

However, observe that:

*A_k requires a **Tail-Head** sequence at times $(k - 1, k)$.*
*A_{k+1} requires a **Tail-Head** sequence at times $(k, k + 1)$.*

*Therefore, for both A_k and A_{k+1} to occur, the sequence from $(k - 1)$ to $(k + 1)$ must be **Tail-Head-Tail**. The probability of this specific sequence occurring is $(1 - p)p(1 - p)$.*

*Let's calculate $P(A_k \cap A_{k+1})$, the probability that both A_k and A_{k+1} happen. This is the probability of the **Tail-Head-Tail** sequence, which is:*

$$P(A_k \cap A_{k+1}) = (1 - p)p(1 - p).$$

Now, since A_k and A_{k+1} are based on overlapping parts of the toss sequence (specifically, the k-th toss), $P(A_k \cap A_{k+1}) \neq P(A_k)P(A_{k+1})$.

Thus, we conclude that:

Events A_k and A_{k+1} are not independent.

Next, let's consider A_k and A_{k+2}.

*For A_k to occur, the sequence from $(k-1, k)$ must be **Tail-Head**. For A_{k+2} to occur, the sequence from $(k+1, k+2)$ must be **Tail-Head**. Since there is no overlap in the outcomes that determine A_k and A_{k+2}, these events are influenced by completely independent tosses.*

Since the coin tosses are independent, the outcomes at times $(k-1, k)$ are independent of the outcomes at $(k+1, k+2)$. Therefore, $P(A_k \cap A_{k+2}) = P(A_k)P(A_{k+2})$.

Exercise 3.7. *A drawer contains two coins. One is an unbiased coin, which when tossed, is equally likely to turn up heads or tails. The other is a biased coin, which will turn up heads with probability p and tails with probability $1-p$. One coin is selected (uniformly) at random from the drawer. Two experiments are performed:*

***(a)** The selected coin is tossed n times. Given that the coin turns up heads k times and tails $n-k$ times, what is the probability that the coin is biased?*

***(b)** The selected coin is tossed repeatedly until it turns up heads k times. Given that the coin is tossed n times in total, what is the probability that the coin is biased?*

Solution 3.7. *Let's tackle each part using Bayes' theorem and probability models.*

***(a)** The selected coin is tossed n times, resulting in k heads and $n-k$ tails. We want to find the probability that the coin is biased, given this outcome. Let's define events for clarity:*

Let B be the event that the coin chosen is biased.
Let U be the event that the coin chosen is unbiased.
Let H_k denote the event of observing k heads in n tosses.

By Bayes' theorem, the probability that the coin is biased given k heads in n tosses is:

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

Since the biased coin produces heads with probability p, if it was chosen, the probability of getting k heads in n tosses is:

$$P(H_k|B) = \binom{n}{k} p^k (1-p)^{n-k}.$$

For the unbiased coin, heads and tails each occur with probability $\frac{1}{2}$. Therefore, if the unbiased coin was chosen, the probability of getting k heads in n tosses is:

$$P(H_k|U) = \binom{n}{k} \left(\frac{1}{2}\right)^n.$$

Since each coin is selected with equal probability, we have $P(B) = P(U) = \frac{1}{2}$. Thus, using the law of total probability:

$$P(H_k) = P(H_k|B) \cdot P(B) + P(H_k|U) \cdot P(U).$$

Substituting the values we have:

$$P(H_k) = \frac{1}{2}\binom{n}{k}p^k(1-p)^{n-k} + \frac{1}{2}\binom{n}{k}\left(\frac{1}{2}\right)^n.$$

Now we can substitute everything back into Bayes' theorem:

$$P(B|H_k) = \frac{\binom{n}{k}p^k(1-p)^{n-k} \cdot \frac{1}{2}}{\frac{1}{2}\binom{n}{k}p^k(1-p)^{n-k} + \frac{1}{2}\binom{n}{k}\left(\frac{1}{2}\right)^n}.$$

Simplifying this expression:

$$P(B|H_k) = \frac{p^k(1-p)^{n-k}}{p^k(1-p)^{n-k} + \left(\frac{1}{2}\right)^n}.$$

***(b)** Now, the coin is tossed repeatedly until it turns up heads k times, with a total of n tosses required to achieve this. We want to find the probability that the coin is biased given these conditions.*

Here, we are interested in the probability that the chosen coin, when tossed, takes n tosses to get k heads. This is a negative binomial setup:

1. If the biased coin is chosen, the probability of getting k heads in n tosses with probability of heads p is:

$$P(H_k|B) = \binom{n-1}{k-1}p^k(1-p)^{n-k}.$$

2. If the unbiased coin is chosen, the probability of getting k heads in n tosses with probability of heads $\frac{1}{2}$ is:

$$P(H_k|U) = \binom{n-1}{k-1}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{n-k} = \binom{n-1}{k-1}\left(\frac{1}{2}\right)^n.$$

Using Bayes' theorem again, we find:

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

With $P(H_k)$ calculated by the law of total probability:

$$P(H_k) = P(H_k|B) \cdot P(B) + P(H_k|U) \cdot P(U),$$

substituting the expressions:

$$P(H_k) = \frac{1}{2}\binom{n-1}{k-1}p^k(1-p)^{n-k} + \frac{1}{2}\binom{n-1}{k-1}\left(\frac{1}{2}\right)^n.$$

Finally,

$$P(B|H_k) = \frac{\binom{n-1}{k-1}p^k(1-p)^{n-k}}{\binom{n-1}{k-1}p^k(1-p)^{n-k} + \binom{n-1}{k-1}\left(\frac{1}{2}\right)^n}.$$

Simplifying,

$$P(B|H_k) = \frac{p^k(1-p)^{n-k}}{p^k(1-p)^{n-k} + \left(\frac{1}{2}\right)^n}.$$

Exercise 3.8. *Fred is giving out samples of dog food. He makes calls door to door, but he leaves a sample (one can) only on those calls for which the door is answered and a dog is in residence. On any call, the probability of the door being answered is $\frac{3}{4}$, and the probability that any household has a dog is $\frac{2}{3}$. Assume that the events "door answered" and "a dog lives here" are independent and also that the outcomes of all calls are independent.*

(a) *Determine the probability that Fred gives away his first sample on his third call.* ***(b)*** *Given that he has given away exactly four samples on his first eight calls, determine the conditional probability that Fred will give away his fifth sample on his eleventh call.*
(c) *Determine the probability that he gives away his second sample on his fifth call.*
(d) *Given that he did not give away his second sample on his second call, determine the conditional probability that he will leave his second sample on his fifth call.*
(e) *We will say that Fred needs a new supply immediately after the call on which he gives away his last can. If he starts out with two cans, determine the probability that he completes at least five calls before he needs a new supply.*

Solution 3.8. *Define the following events:*

Let A denote the event the door is answered.
Let B denote the event a dog lives in the house.

Since events A and B are independent, the probability that Fred gives away a sample (which requires both events A and B to happen) on any given call is given by:

$$P(\textit{sample given}) = P(A \cap B) = P(A) \cdot P(B) = \frac{3}{4} \times \frac{2}{3} = \frac{1}{2}$$

We can now proceed to solve each part of the problem.

(a) *Determine the probability that Fred gives away his first sample on his third call.*

This situation describes a classic ***Geometric Distribution*** *because we're looking for the probability that the first successful outcome (i.e., giving away a sample) happens on the third trial. If we let $p = \frac{1}{2}$ be the probability of success, then the probability of the first success occurring on the k-th trial is given by:*

$$P(\textit{first success on kth call}) = (1-p)^{k-1}p$$

For this problem, $k = 3$, so:

$$P(\text{first sample on third call}) = (1 - \frac{1}{2})^{3-1} \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{8}$$

(b) *Given that he has given away exactly four samples on his first eight calls, determine the conditional probability that Fred will give away his fifth sample on his eleventh call.*

Here, we are given that exactly four samples were given in the first eight calls. The probability of giving away four samples in eight calls follows a ***Binomial Distribution*** *with $n = 8$ trials and success probability $p = \frac{1}{2}$. The probability of giving the fifth sample on the eleventh call can be analyzed by considering two conditions:*

1. The next two calls (the ninth and tenth) result in no sample being given, which has probability $(1 - p)^2$.
2. The eleventh call results in a sample being given, with probability p.

Thus, the conditional probability is:

$$P(\text{fifth sample on 11th call} \mid \text{four samples in first eight calls}) = (1-p)^2 \cdot p = \left(\frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{8}$$

(c) *Determine the probability that he gives away his second sample on his fifth call.*

This follows a ***Negative Binomial Distribution****, as we are interested in the probability of the second success occurring on the fifth call. For a negative binomial event, the probability of the r-th success on the k-th trial is:*

$$P(\text{second success on fifth call}) = \binom{4}{1} p^2 (1-p)^3$$

where $p = \frac{1}{2}$. Thus,

$$P(\text{second success on fifth call}) = \binom{4}{1} \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^3 = 4 \cdot \frac{1}{32} = \frac{1}{8}$$

(d) *Given that he did not give away his second sample on his second call, determine the conditional probability that he will leave his second sample on his fifth call.*

Given that the second sample was not given on the second call, we want the probability of the second success on the fifth call. This conditional probability is based on trials up to the fifth call with one success among the first four trials, resulting in a similar calculation to part (c):

$$P(\text{second sample on fifth call} \mid \text{one success in four trials}) = \frac{1}{8}$$

(e) *We will say that Fred needs a new supply immediately after the call on which he gives away his last can. If he starts out with two cans, determine the probability that he completes at least five calls before he needs a new supply.*

To find this probability, Fred needs exactly two successes in the first five calls (so he gives away both cans by the fifth call). Using a binomial distribution with parameters $n = 4$ and $p = \frac{1}{2}$, we calculate the probability of having exactly two successes in the first four calls, which would mean he finishes his supply in four calls. Thus, we need to find the probability that he does not finish his supply within the first four calls, which would imply he still has at least one can left by the fifth call.

Let X be the number of samples given in the first four calls. We want $P(X < 2)$, which is the probability of giving away fewer than two samples in four calls.

Using the binomial formula:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

we calculate $P(X = 0)$ and $P(X = 1)$.

1. For $X = 0$:

$$P(X = 0) = \binom{4}{0} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4 = 1 \cdot \frac{1}{16} = \frac{1}{16}$$

2. For $X = 1$:

$$P(X = 1) = \binom{4}{1} \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^3 = 4 \cdot \frac{1}{16} = \frac{4}{16} = \frac{1}{4}$$

Thus, the probability that he has fewer than two successes in the first four calls (i.e., that he does not run out of samples) is:

$$P(X < 2) = P(X = 0) + P(X = 1) = \frac{1}{16} + \frac{1}{4} = \frac{1}{16} + \frac{4}{16} = \frac{5}{16}$$

Exercise 3.9. *Let $A, B, A_1, A_2, \ldots$ be events. Suppose that for each k, we have*

$$A_k \subseteq A_{k+1},$$

and that A_k is independent of B for all $k \geq 1$. If we define $A = \bigcup_{k \in \mathbb{N}} A_k$, show that B is independent of A.

Solution 3.9. *The key to solving this problem lies in understanding how independence works for increasing events. Let us start with the definitions and properties we'll use:*

1. Since $A_k \subseteq A_{k+1}$ for all k, the sequence $\{A_k\}_{k \geq 1}$ is an increasing sequence of events. Therefore, $A = \bigcup_{k=1}^{\infty} A_k$ represents the limit of this increasing sequence, formally written as:

$$A = \lim_{k \to \infty} A_k.$$

2. Given that each A_k is independent of B, we know:

$$P(A_k \cap B) = P(A_k)P(B) \quad \textit{for each } k \geq 1.$$

3. To show that A is independent of B, we need to prove:

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

Now, let's carefully calculate $P(A \cap B)$ and $P(A)$ using the properties of limits of probabilities in increasing sequences.

Step 1: Calculating $P(A \cap B)$

Since $A_k \subseteq A_{k+1}$, the sequence $\{A_k \cap B\}_{k \geq 1}$ is also an increasing sequence of events. Therefore, we can use the continuity property of probability for increasing sequences:

$$P(A \cap B) = P\left(\lim_{k \to \infty} (A_k \cap B)\right) = \lim_{k \to \infty} P(A_k \cap B).$$

From the independence of A_k and B, we know that $P(A_k \cap B) = P(A_k)P(B)$ for each k. Substituting this in, we get:

$$P(A \cap B) = \lim_{k \to \infty} P(A_k)P(B).$$

Since $P(B)$ is a constant with respect to k, we can factor it out of the limit:

$$P(A \cap B) = P(B) \cdot \lim_{k \to \infty} P(A_k).$$

Step 2: Calculating $P(A)$

Again, because A_k is an increasing sequence of events with $A = \bigcup_{k=1}^{\infty} A_k$, we use the continuity property of probability:

$$P(A) = P\left(\lim_{k \to \infty} A_k\right) = \lim_{k \to \infty} P(A_k).$$

Step 3: Verifying Independence

Now, substituting $P(A) = \lim_{k \to \infty} P(A_k)$ from Step 2 into our expression from Step 1, we obtain:

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

Exercise 3.10. *Consider pairwise disjoint events B_1, B_2, B_3, and C with probabilities*

- $P(B_1) = P(B_2) = P(B_3) = p$
- $P(C) = q$,

where $3p + q \leq 1$. Suppose $p = -q + \sqrt{q}$. Prove that the events $B_1 \cup C$, $B_2 \cup C$, and $B_3 \cup C$ are pairwise independent. Additionally, determine whether there exist values $p > 0$ and $q > 0$ such that these three events are independent.

Solution 3.10. *First, let's analyze the pairwise independence of the events $B_1 \cup C$, $B_2 \cup C$, and $B_3 \cup C$.*

Since B_1, B_2, B_3, and C are pairwise disjoint, we have:

$$P(B_i \cap B_j) = 0, \quad \text{for any } i \neq j.$$

For each B_i, $i = 1, 2, 3$, the probability of the event $B_i \cup C$ is given by:

$$P(B_i \cup C) = P(B_i) + P(C) = p + q.$$

The intersection of $B_i \cup C$ and $B_j \cup C$ (for $i \neq j$) can be expressed as:

$$(B_i \cup C) \cap (B_j \cup C) = (B_i \cap B_j) \cup (B_i \cap C) \cup (B_j \cap C) \cup (C \cap C).$$

Given that B_i and B_j are disjoint (so $B_i \cap B_j = \emptyset$), this simplifies to:

$$(B_i \cup C) \cap (B_j \cup C) = C.$$

Thus, we have:

$$P((B_i \cup C) \cap (B_j \cup C)) = P(C) = q.$$

To verify pairwise independence of $B_1 \cup C$, $B_2 \cup C$, and $B_3 \cup C$, we need to check if:

$$P((B_i \cup C) \cap (B_j \cup C)) = P(B_i \cup C) \cdot P(B_j \cup C).$$

Substituting from above, we have:

$$q = (p + q)^2.$$

Expanding the right side:

$$q = p^2 + 2pq + q^2.$$

Rearrange terms to form a quadratic equation in p:

$$p^2 + (2q)p + (q^2 - q) = 0.$$

Solving this quadratic equation for p, we find:

$$p = -q + \sqrt{q}.$$

Thus, given $p = -q + \sqrt{q}$, we have shown that $B_1 \cup C$, $B_2 \cup C$, and $B_3 \cup C$ are pairwise independent.

For mutual independence, we need:

$$P((B_1 \cup C) \cap (B_2 \cup C) \cap (B_3 \cup C)) = P(B_1 \cup C) \cdot P(B_2 \cup C) \cdot P(B_3 \cup C).$$

The left side represents the probability of the intersection of all three events:

$$(B_1 \cup C) \cap (B_2 \cup C) \cap (B_3 \cup C) = C,$$

since B_1, B_2, B_3 are pairwise disjoint. Thus,

$$P((B_1 \cup C) \cap (B_2 \cup C) \cap (B_3 \cup C)) = P(C) = q.$$

The right side is:

$$P(B_1 \cup C) \cdot P(B_2 \cup C) \cdot P(B_3 \cup C) = (p + q)^3.$$

For mutual independence, we need $q = (p + q)^3$.

Substituting $p = -q + \sqrt{q}$ into this equation does not generally satisfy equality for all $p > 0$ and $q > 0$.

Therefore, while the events $B_1 \cup C$, $B_2 \cup C$, and $B_3 \cup C$ are pairwise independent, they are not mutually independent for any choice of $p > 0$ and $q > 0$.

3.3 Borel-Cantelli Lemmas

Let's imagine we're playing a game of chance over and over again, like flipping a coin or drawing cards, and we want to understand if certain events will happen *infinitely often* or just a *finite* number of times. In probability, we often ask questions like, *If I keep playing this game, will I see a particular outcome repeatedly?*

Now, that's where the Borel-Cantelli Lemmas come into play! These lemmas help us decide, based on probabilities, if an event is bound to happen over and over or only occasionally.

The Borel-Cantelli Lemmas are two parts that answer different versions of our question.

Part 1 tells us that if the sum of probabilities of all events A_n is *finite*, then almost surely only a finite number of these events will occur. In simpler terms, if the probabilities of each event happening are so small that their total barely adds up to anything, then we shouldn't expect to see these events happening infinitely often. Mathematically, if

$$\sum_{n=1}^{\infty} P(A_n) < \infty$$

then the probability that infinitely many of the A_n happen is zero. In notation:

$$P\left(\limsup_{n\to\infty} A_n\right) = 0,$$

where $\limsup_{n\to\infty} A_n$ represents the event that infinitely many of the A_n occur.

Now, what if the probabilities of the events don't just add up to something finite but actually keep adding up indefinitely? If the events are *independent*, then *Part 2* of the Borel-Cantelli Lemma kicks in and tells us that, in this case, infinitely many of these events will indeed occur. That is, if

$$\sum_{n=1}^{\infty} P(A_n) = \infty$$

and the events A_n are independent, then:

$$P\left(\limsup_{n\to\infty} A_n\right) = 1.$$

This means that as we keep going, we're *guaranteed* to see infinitely many occurrences of the events A_n.

If you're flipping a fair coin forever, the event *heads on flip n* is independent for each n and has a fixed probability of $\frac{1}{2}$. According to *Part 2* of Borel-Cantelli, because the total sum of probabilities grows infinitely (since $\sum \frac{1}{2}$ diverges), we'll keep seeing heads infinitely often as we flip the coin forever.

Lemma 3.1. ***First Borel-Cantelli Lemma.*** *Suppose we have a sequence of events, $\{A_n\}_{n=1}^{\infty}$, and their probabilities $P(A_n)$ add up to something finite:*

$$\sum_{n=1}^{\infty} P(A_n) < \infty.$$

Then the first Borel-Cantelli lemma tells us that in this case, only a finite number of the events A_n will occur with probability 1, or ***almost surely****. This means that as n grows, we reach a point where none of the remaining A_n events happen, essentially running out of events that can occur.*

Lemma 3.2. ***Second Borel-Cantelli Lemma.*** *Now, if we change our setup slightly and assume that the sequence $\{A_n\}_{n=1}^{\infty}$ consists of* ***independent*** *events, and if the probabilities $P(A_n)$ now add up to infinity:*

$$\sum_{n=1}^{\infty} P(A_n) = \infty,$$

then the second Borel-Cantelli lemma says that infinitely many of the A_n events will occur almost surely. Independence is key here. Without it, this conclusion might not hold.

The statement $\{A_n \text{ i.o.}\}$, meaning $\{A_n \text{ occurs infinitely often}\}$, represents the set of all outcomes $\omega \in \Omega$ that belong to infinitely many of the events A_n. We define this as follows:

$$\{A_n \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \equiv B_n.$$

Here, B_n is the event that at least one of the events $A_n, A_{n+1}, A_{n+2}, \ldots$ occurs. Thus, $\{A_n \text{ i.o.}\}$ is the event that for any positive integer n, there exists some $m \geq n$ such that A_m happens. In other words, we are always *catching* one of the A_m events, no matter how far out we go in the sequence.

To understand the event that A_n happens **finitely often** (or $\{A_n \text{ f.o.}\}$), we can take the complement of the event $\{A_n \text{ i.o.}\}$:

$$\{A_n \text{ f.o.}\} = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m^c,$$

where A_m^c denotes the complement of A_m, i.e., the event that A_m does not occur.

To prove the Borel-Cantelli lemmas, we need the following foundational lemma:

Lemma 3.3. *Suppose $\sum_{i=1}^{\infty} p_i = \infty$. Then,*

$$\lim_{n\to\infty} \prod_{i=1}^{n} (1 - p_i) = 0.$$

Proof. We begin by observing that the natural logarithm of each term satisfies an upper bound:

$$\ln(1 - p_i) \leq -p_i.$$

Using this, we can express the product $\prod_{i=1}^{n}(1-p_i)$ in terms of exponentials:

$$\prod_{i=1}^{n}(1-p_i) = \prod_{i=1}^{n} e^{\ln(1-p_i)} \leq \prod_{i=1}^{n} e^{-p_i} = e^{-\sum_{i=1}^{n} p_i}.$$

Taking the limit as $n \to \infty$, we find

$$\lim_{n\to\infty} \prod_{i=1}^{n}(1-p_i) \leq \lim_{n\to\infty} e^{-\sum_{i=1}^{n} p_i}.$$

Since $\sum_{i=1}^{\infty} p_i = \infty$, the partial sums $\sum_{i=1}^{n} p_i$ tend to infinity as $n \to \infty$. Therefore,

$$\lim_{n\to\infty} e^{-\sum_{i=1}^{n} p_i} = 0.$$

Thus, we conclude that

$$\lim_{n\to\infty} \prod_{i=1}^{n}(1-p_i) = 0.$$

□

We now proceed towards proving the Borel-Cantelli lemmas.

Proof. **First Borel-Cantelli Lemma.**

We start by noting that the sum of probabilities over the events A_n, given by $\sum_{n=1}^{\infty} P(A_n)$, converges. This means that:

$$\sum_{m=n}^{\infty} P(A_m) \to 0 \quad \text{as} \quad n \to \infty.$$

This result follows directly from the convergence of the series $\sum_{n=1}^{\infty} P(A_n)$, implying that as we go further in the sequence, the cumulative probability from any point n onward must approach zero.
Now, let us define a sequence of events B_n as:

$$B_n = \bigcup_{m=n}^{\infty} A_m,$$

which represents the occurrence of at least one of the events A_m for $m \geq n$. Notice that these sets B_n form a decreasing sequence since:

$$B_{n+1} \subset B_n.$$

By the continuity of probability for decreasing events, we can write:

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} P(B_n).$$

Since $B_n = \bigcup_{m=n}^{\infty} A_m$, we have:

$$P(B_n) \leq \sum_{m=n}^{\infty} P(A_m).$$

Taking the limit as $n \to \infty$, we find:

$$\lim_{n\to\infty} P(B_n) \leq \lim_{n\to\infty} \sum_{m=n}^{\infty} P(A_m) = 0.$$

Therefore:

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = 0.$$

Second Borel-Cantelli Lemma.

To understand the probability of an event A_n occurring only finitely often, we begin by defining the event that A_n occurs finitely often (denoted as $\{A_n \text{ f.o.}\}$) as follows:

$$\{A_n \text{ f.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i^c$$

where A_i^c denotes the complement of A_i, representing the event that A_i does not happen. This setup allows us to examine the probability that after some point n, none of the events A_i occur. We then proceed to calculate this probability using a series of bounds and properties of probability.

First, applying the *union bound* (which states that the probability of a union of events is less than or equal to the sum of the probabilities of each event), we obtain:

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i^c\right) \leq \sum_{n=1}^{\infty} P\left(\bigcup_{i=n}^{\infty} A_i^c\right)$$

Next, by the *continuity of probability*, we rewrite the probability of the infinite union as the limit of finite unions:

$$= \sum_{n=1}^{\infty} \lim_{m\to\infty} P\left(\bigcup_{i=n}^{m} A_i^c\right)$$

Given the *independence* of the events A_i, we can further simplify each term in this sum by multiplying the probabilities of each A_i^c:

$$= \sum_{n=1}^{\infty} \prod_{i=n}^{\infty} P(A_i^c)$$

According to **Lemma 3.3**, this product approaches zero as $n \to \infty$, yielding:

$$= 0$$

Since $P\left(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i^c\right) \geq 0$, we conclude that:

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i^c\right) = 0$$

This result implies that the probability of the event A_n occurring infinitely often is equal to 1, meaning that, with probability 1, A_n will occur infinitely often.

$\square$

Example 3.1. *Let's consider an experiment in which we toss a coin repeatedly and independently. Let the probability of obtaining a head on the n-th toss be denoted as $P(H_n)$, and similarly $P(T_n)$ for tails.*

1. Case 1: Suppose $P(H_n) = \frac{1}{n}$ ***for*** $n \geq 1$.

In this case, we can sum the probabilities over all tosses:

$$\sum_{n=1}^{\infty} P(H_n) = \sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

This series diverges, meaning it adds up to infinity. Now, by the ***Second Borel-Cantelli Lemma****, we conclude that almost surely, there will be infinitely many heads in the sequence of tosses.*

This result might initially seem counterintuitive, as the probability of getting a head decreases with each toss — it becomes extremely small as n grows. However, the decay rate $\frac{1}{n}$ is not fast enough to prevent heads from occurring infinitely often. In fact, no matter how large we make n, there will almost surely be a head occurring somewhere after the n-th toss.

2. Case 2: Suppose $P(H_n) = \frac{1}{n^2}$.

Now, let's examine what happens if the probability of getting a head on the n-th toss decays faster, specifically as $\frac{1}{n^2}$:

$$\sum_{n=1}^{\infty} P(H_n) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

This series converges, meaning it sums to a finite value. By the ***First Borel-Cantelli Lemma****, we conclude that almost surely, only finitely many heads will occur.*

In this scenario, the probability of obtaining a head decreases so rapidly that, after a certain finite number of tosses, the likelihood of obtaining further heads becomes negligible. The decay rate $\frac{1}{n^2}$ is fast enough that, beyond some large n, we can almost be certain that no more heads will appear.

Exercise 3.11. *Consider a monkey sitting in front of a computer and randomly pressing keys on the keyboard. We want to demonstrate that the complete monologue by Shakespeare, which begins with "All the world's a stage", will eventually appear on the screen with a probability of 1. This conclusion may seem surprising, as the monkey is not recognized for its literary talent.*

Solution 3.11. *To tackle this problem, we can set up a probability model with a few reasonable assumptions. We will assume the following:*

1. The monkey chooses each character from the keyboard uniformly at random.
2. Each key stroke made by the monkey is independent of previous strokes.
3. The keyboard consists of a finite set of characters, which includes letters, spaces, and punctuation marks. Let us denote this set as $\mathcal{K}$.

Let n be the total number of characters in the Shakespeare monologue we are interested in.

Each time the monkey types a key, it selects a character from $\mathcal{K}$, which contains m characters (including letters, spaces, and punctuation).

The probability that the monkey correctly types the first character of the monologue is given by:

$$P(\textit{first character}) = \frac{1}{m}$$

Similarly, the probability of typing the second character correctly after the first is:

$$P(\textit{second character}) = \frac{1}{m}$$

Continuing this reasoning, the probability of typing the entire monologue correctly in any n consecutive keystrokes is:

$$P(\textit{monologue}) = \left(\frac{1}{m}\right)^n$$

Now, consider the total number of keystrokes the monkey can make. If the monkey types continuously, the number of keystrokes approaches infinity as time goes on. To find the probability of the monologue appearing at least once in this infinite series of keystrokes, we use the complement probability.

Let A be the event that the monologue appears at least once. The complement of A, denoted as A^c, is the event that the monologue does not appear in k keystrokes. The probability of not typing the monologue in k trials is:

$$P(A^c) = 1 - P(\textit{monologue})^k = 1 - \left(1 - \frac{1}{m^n}\right)^k$$

As k approaches infinity, the expression $\left(1 - \frac{1}{m^n}\right)^k$ converges to 0, since $\frac{1}{m^n}$ is a small positive number. Thus:

$$\lim_{k\to\infty} P(A^c) = 0$$

Consequently, we find that:

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

Exercise 3.12. *Let us consider a sequence of events A_n for $n \geq 1$ such that the probability of each event tends to zero as n approaches infinity, denoted as $P(A_n) \to 0$ as $n \to \infty$. Furthermore, we know that the sum of the probabilities of the intersections of the complements of these events with the subsequent events is finite:*

$$\sum_{n=1}^{\infty} P(A_n^c \cap A_{n+1}) < \infty.$$

We aim to demonstrate that, almost surely, only finitely many of the events A_n will occur.

Solution 3.12. *To establish this, we utilize the concept of the Borel-Cantelli lemma, which provides insight into the occurrence of events based on their probabilities.*

Firstly, we denote $B_n = A_n^c \cap A_{n+1}$. The significance of this intersection is that B_n represents the scenario where A_n does not occur while A_{n+1} does. Hence, if we have a finite sum of probabilities:

$$\sum_{n=1}^{\infty} P(B_n) < \infty,$$

the Borel-Cantelli lemma informs us that the probability that infinitely many of the events B_n occur is zero. In other words, almost surely, there will be only finitely many n for which A_n does not happen followed by A_{n+1} happening.

Now, we define the event $C = \{\text{infinitely many } A_n \text{ occur}\}$. The complement of C, denoted C^c, represents the scenario where only finitely many of the A_n occur.

From our previous work, we conclude that if infinitely many B_n occur, then:

$$P(C) = P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} B_n\right) = 0.$$

Consequently, we deduce that:

$$P(C^c) = 1.$$

Thus, we have shown that almost surely, only finitely many of the events A_n will occur.

Exercise 3.13. *On a certain day, Alice decides that she will start looking for a potential life partner on an online dating portal. She decides that everyday, she will pick a guy uniformly at random from among the male members of the dating portal, and go out on a date with him. What Alice does not know, is that her neighbor Bob is interested in dating her. Being of a shy disposition, Bob decides that he will not ask Alice out himself. Instead, he decides that he will go out on a date with Alice only on the days that Alice happens to pick him from the dating portal, of which he is already a member. For the first two parts, assume that 50 new male members and 40 new female members join the dating portal everyday.*

(a) What is the probability that Alice and Bob would have a date on the nth day? Do you think Bob and Alice would eventually stop meeting? Justify your answer, clearly stating any additional assumptions.

(b) Now suppose that Bob also picks a girl uniformly at random everyday, from among the female members of the portal, and that Alice behaves exactly as before. Assume also that Bob and Alice will meet on a given day if and only if they both happen to pick each other. In this case, do you think Bob and Alice would eventually stop meeting?

(c) For this part, suppose that Alice and Bob behave as in part (a), i.e.,Alice picks a guy uniformly at random, but Bob is only interested in dating Alice. However, the number of male members in the portal increases by 1 percent everyday. Do you think Bob and Alice would eventually stop meeting?

Solution 3.13. *(a) To determine the probability that Alice and Bob will have a date on the nth day, we start by defining the total number of male members in the dating portal.*

Initially, let M_n denote the total number of male members on the nth day. We can express this as:

$$M_n = M_0 + 50n$$

where M_0 is the number of male members present at the start (day 0).

Now, since Bob is one of these male members, the probability that Alice randomly selects Bob on the nth day is given by:

$$P(Alice\ picks\ Bob\ on\ day\ n) = \frac{1}{M_n} = \frac{1}{M_0 + 50n}$$

For Bob, the chance of Alice selecting him must be matched by the condition that Alice happens to choose him out of the total male members on that day. Hence, the probability that they have a date on the nth day is:

$$P(Alice\ and\ Bob\ have\ a\ date\ on\ day\ n) = \frac{1}{M_0 + 50n}$$

Now, regarding whether Alice and Bob will eventually stop meeting, we observe that as n increases, M_n continues to grow because it is increasing linearly with time.

Thus, the probability $P(Alice\ and\ Bob\ have\ a\ date\ on\ day\ n)$ approaches zero as n approaches infinity:

$$\lim_{n\to\infty} P(Alice\ and\ Bob\ have\ a\ date\ on\ day\ n) = 0$$

This suggests that, as the number of male members increases indefinitely, Alice and Bob will eventually stop meeting.

(b) In this scenario, let us consider the case where Bob also randomly selects a female member each day from the pool of female members, while Alice continues her previous behavior. They will meet only if both select each other.

Let F_n be the total number of female members on the nth day, given by:

$$F_n = F_0 + 40n$$

The probability that Alice picks Bob remains the same as before:

$$P(Alice\ picks\ Bob) = \frac{1}{M_n} = \frac{1}{M_0 + 50n}$$

Now for Bob to pick Alice, the probability is:

$$P(Bob\ picks\ Alice) = \frac{1}{F_n} = \frac{1}{F_0 + 40n}$$

Thus, the probability that they will meet on the nth day is:

$$P(Alice\ and\ Bob\ meet) = P(Alice\ picks\ Bob) \times P(Bob\ picks\ Alice) = \frac{1}{(M_0 + 50n)(F_0 + 40n)}$$

Similar to the first part, as n increases, both M_n and F_n grow, leading this probability to approach zero:

$$\lim_{n\to\infty} P(Alice\ and\ Bob\ meet) = 0$$

Thus, we can conclude that Alice and Bob will also eventually stop meeting in this situation.

(c) In this case, we revert to the situation described in part (a) where Alice randomly selects a male member uniformly, while Bob is solely interested in dating Alice. However, the twist is that the number of male members increases by 1 percent daily. This means that:

$$M_n = M_0(1.01)^n$$

Consequently, the probability that Alice picks Bob remains:

$$P(Alice\ picks\ Bob\ on\ day\ n) = \frac{1}{M_n} = \frac{1}{M_0(1.01)^n}$$

As n increases, M_n grows exponentially, leading to:

$$\lim_{n\to\infty} P(Alice\ picks\ Bob\ on\ day\ n) = 0$$

indicating that the probability of Alice and Bob going on a date diminishes over time. Therefore, they will ultimately stop meeting, as the number of male members is growing faster than linearly, reinforcing the conclusion that Alice and Bob will cease to meet over time.

Exercise 3.14. *Let $S_n : n \geq 0$ be a simple random walk defined such that it moves to the right with probability p at each step. We begin with $S_0 = 0$. We denote $X_n = S_n - S_{n-1}$, which represents the change in position at step n. Show that:*
(a) $S_n = 0$ i.o. is not a tail event of the sequence X_n
(b) $P(S_n = 0 \text{ i.o.}) = 0$ if $p \neq \frac{1}{2}$

Solution 3.14. *A tail event is an event whose occurrence or non-occurrence is independent of the outcomes of any finite number of preceding steps in the sequence.*

The event $S_n = 0$ i.o. means that the random walk returns to the origin infinitely often. In contrast, the sequence $\{X_n\}$ comprises the individual steps of the random walk, which can be thought of as either moving right with probability p or left with probability $1 - p$.

If we consider the finite sum $S_n = \sum_{i=1}^{n} X_i$, we see that the occurrence of $S_n = 0$ i.o. depends on the entire history of the steps taken, not merely the individual steps represented by X_n. Therefore, the event $S_n = 0$ i.o. cannot be determined solely by the values of X_n and thus is not a tail event.

When $p \neq \frac{1}{2}$, the random walk has a bias in either direction (either more likely to move right if $p > \frac{1}{2}$ or more likely to move left if $p < \frac{1}{2}$). In this scenario, the random walk will drift away from the origin over time.

By the Law of Large Numbers, as n becomes large, the average position of the random walk approaches the expected value. Specifically, since $E[X_n] = p - (1 - p) = 2p - 1$, we find that:

$$E[S_n] = n(2p - 1)$$

If $p > \frac{1}{2}$, then $E[S_n] \to +\infty$ as $n \to \infty$, and if $p < \frac{1}{2}$, then $E[S_n] \to -\infty$ as $n \to \infty$. Thus, the walk will not return to 0 infinitely often, leading to:

$$P(S_n = 0 \text{ i.o.}) = 0 \quad \text{for } p \neq \frac{1}{2}.$$

Problem Set 1 - Review

Chapter 4

Random Variables

Chapter 5

Transformation of Random Variables

Chapter 6

Expectation and Variance

Chapter 7

Generating Functions and Inequalities

Chapter 8

Limit Theorems

Bibliography

[1] Krishna Jagannath. *Probability Foundations for Electrical Engineers*. IIT Madras, 2015.