Measure Theoretic Probability

Prerequisites: Real Analysis, Classical Probability

Sagar Udasi

November 3, 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 $A_1, A_2, A_3, \dots$ 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:
$$A_1 = \{1\}, \quad A_2 = \{2\}, \quad A_3 = \{3\}, \quad \ldots$$
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 $\{1\}, \{2\}, \{3\}, \ldots$, 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 $0 \leq a \leq b \leq 1$. 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:

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

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:

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

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

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:*

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

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

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

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:

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

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:

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

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 $A = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$, 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 $(0,1]$. $\square$

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.

$\square$

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 $\mathcal{F}$.

□

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:

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

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:

$$B_i \cap B_j = \varnothing \quad \textit{for all } i \neq j.$$

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 $F_1 = \{\varnothing, \Omega, A, A^c\}$ and $F_2 = \{\varnothing, \Omega, B, B^c\}$. 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 $B = D$:

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

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:

$$\begin{aligned} P(B) &= \frac{3}{4} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{5}{6} \\ &= \frac{3}{24} + \frac{5}{24} = \frac{8}{24} = \frac{1}{3} \end{aligned}$$

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:

$$P(\textit{first sample on third call}) = (1 - \frac{1}{2})^{3-1} \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{8}$$

(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:

$$P(\textit{fifth sample on 11th call} \mid \textit{four samples in first eight calls}) = (1-p)^2 \cdot p = \left(\frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{8}$$

(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:*

$$P(\textit{second success on fifth call}) = \binom{4}{1} p^2 (1-p)^3$$

where $p = \frac{1}{2}$. Thus,

$$P(\textit{second success on fifth call}) = \binom{4}{1} \cdot \left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^3 = 4 \cdot \frac{1}{32} = \frac{1}{8}$$

(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):

$$P(\textit{second sample on fifth call} \mid \textit{one success in four trials}) = \frac{1}{8}$$

(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:

$$P(A_k \cap B) = P(A_k)P(B) \quad \text{for each } k \geq 1.$$

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.

□

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 $C = \{$infinitely many A_n 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:

$$P(Alice \text{ picks Bob on day } n) = \frac{1}{M_n} = \frac{1}{M_0(1.01)^n}$$

As n increases, M_n grows exponentially, leading to:

$$\lim_{n\to\infty} P(Alice \text{ picks Bob on day } n) = 0$$

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

The exploration of random variables arises from the observation that, in various situations, we may not be concerned with the specific basic outcome of a random experiment. Instead, we are often interested in some numerical function derived from that outcome. For instance, consider an experiment where a coin is tossed ten times. Rather than focusing on the exact sequence of heads and tails that appears—such as heads, tails, heads, heads, tails, etc.—the experimenter might be primarily interested in the total count of heads obtained.

This shift in focus allows us to encapsulate the randomness of the experiment in a more manageable and meaningful way, as we transform the complexity of individual outcomes into a single number that conveys essential information about the experiment's result.

4.1 Introduction to Random Variables

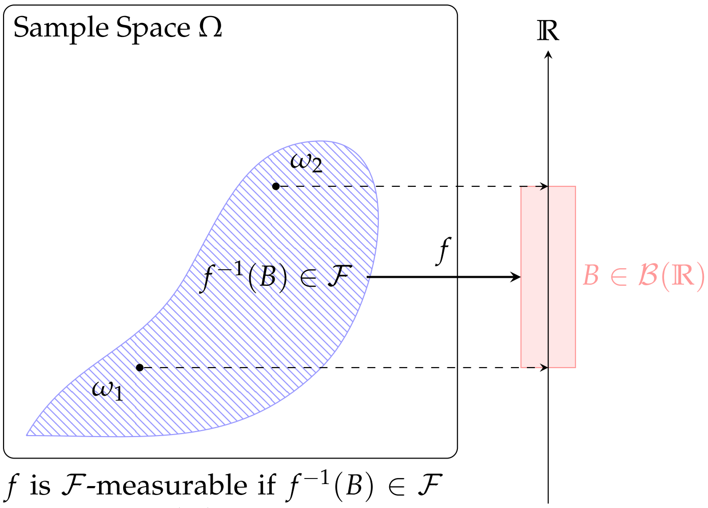
The term **random variable** can be misleading because it suggests that the variable itself is random, or that it varies in a typical sense. In fact, a random variable X is better understood as a function that maps elements from the sample space Ω to the real numbers $\mathbb{R}$. The term *random* refers to the inherent uncertainty in selecting an element ω from the sample space Ω. Once we fix an elementary outcome ω, the random variable assigns a specific real value, denoted $X(\omega)$.

It is crucial to distinguish between the probability measure, which applies to subsets of the sample space (known as events), and the random variable, which is tied to each individual outcome ω. Not all subsets of the sample space qualify as events, and similarly, not every function from Ω to $\mathbb{R}$ is classified as a random variable. In particular, a random variable is defined as an $\mathcal{F}$**-measurable function**, as elaborated below.

Definition 4.1. *Consider a measurable space $(\Omega, \mathcal{F})$. A function $f : \Omega \to \mathbb{R}$ is termed an $\mathcal{F}$-**measurable function** if the pre-image of every Borel set is an $\mathcal{F}$**-measurable** subset of Ω.*

To clarify, the pre-image of a Borel set B under the function f is defined as:

$$f^{-1}(B) = \{\omega \in \Omega \mid f(\omega) \in B\}$$



f is $\mathcal{F}$-measurable if $f^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R})$

Definition 4.2. *Consider a probability space $(\Omega, \mathcal{F}, P)$, where Ω represents the sample space, $\mathcal{F}$ is a σ-algebra of events, and P is a probability measure. A random variable X is defined as a function that maps elements from Ω to the real numbers $\mathbb{R}$, such that X is measurable with respect to $\mathcal{F}$.*

To elaborate, this means that for any Borel set B (which belongs to the collection of Borel sets $\mathcal{B}(\mathbb{R})$), the pre-image of B under the random variable X must be an event in the σ-algebra $\mathcal{F}$. In a visual sense, you can think of X as a mapping that takes each outcome ω from the sample space Ω and assigns it a value in the real line $\mathbb{R}$.

Now, the set defined as $\{\omega \in \Omega \mid X(\omega) \in B\}$ represents an event in $\mathcal{F}$ for every Borel set B. This set corresponds to the outcomes for which the random variable X falls within the Borel set B. As a result, since each such event has a probability associated with it, we can define what is known as the probability law of the random variable X. This law encapsulates how probabilities are distributed across the possible values that X can take.

Definition 4.3. *The* ***probability law****, denoted as P_X, is defined as a function that maps Borel sets to probabilities, specifically:*

$$P_X : \mathcal{B}(\mathbb{R}) \to [0, 1]$$

This means that for any Borel set B in $\mathbb{R}$, the probability law is given by:

$$P_X(B) = P(\{\omega \in \Omega \mid X(\omega) \in B\})$$

Here, P is the probability measure applied to the set of outcomes ω in the sample space Ω for which the random variable X falls within the set B.

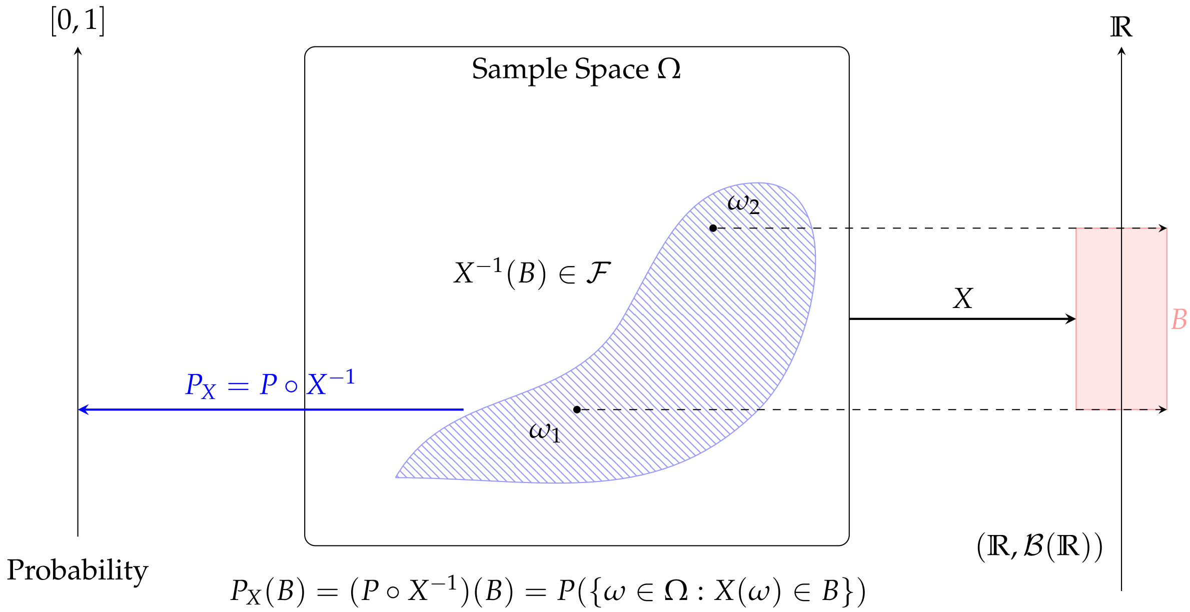
To clarify further, we can think of P_X as a composition of two functions: the probability measure P and the inverse image of X, denoted as X^{-1}. Mathematically, we express this relationship as:

$$P_X(\cdot) = P \circ X^{-1}(\cdot)$$

Notice that we are not saying X^{-1} is an inverse function i.e., the random variable function X is invertible. This is not the case. That's just the notation for the set of elements of the sample space that map to the Borel set B. So, X^{-1} is the pre-image of B. Since X is an $\mathcal{F}$-measurable function, X^{-1} is the element of the Borel σ-algebra of Ω.

This relationship reveals that the probability law completely determines the statistical properties of the random variable X. In other words, it tells us the likelihood of X taking values in any given Borel set B.

To visualize this, consider that P represents the mapping from an event E to its associated probability. In contrast, P_X maps a Borel set B to the probability space, effectively combining the function P with the inverse image X^{-1}. This process illustrates how the behavior of the random variable is encapsulated by its probability law.



Theorem 4.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X be a real-valued random variable. Then, the probability law P_X of X is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

Proof. We need to show that P_X is indeed a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

1. **Non-negativity:** For any Borel set B, since P is a probability measure, we have:

$$P_X(B) = P(X^{-1}(B)) \geq 0.$$

2. **Normalization:** The total measure of the whole space must equal 1. Since X can take any value in $\mathbb{R}$, we have:

$$P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1.$$

3. **Countable Additivity:** For any countable collection of disjoint Borel sets $B_1, B_2, \ldots$ (i.e., $B_i \cap B_j = \emptyset$ for $i \neq j$), we have:

$$P_X\left(\bigcup_{i=1}^{\infty} B_i\right) = P(X^{-1}(\bigcup_{i=1}^{\infty} B_i)) = P\left(\bigcup_{i=1}^{\infty}\{\omega \in \Omega : X(\omega) \in B_i\}\right).$$

By the countable additivity of P, this can be rewritten as:

$$P\left(\bigcup_{i=1}^{\infty}\{\omega \in \Omega : X(\omega) \in B_i\}\right) = \sum_{i=1}^{\infty} P\left(\{\omega \in \Omega : X(\omega) \in B_i\}\right) = \sum_{i=1}^{\infty} P_X(B_i).$$

Since all three properties of a probability measure hold, we conclude that P_X is indeed a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. $\square$

Note: If Ω is countable set, then $\mathcal{F}$ can be taken as 2^{Ω}. Typical example is the *discrete probability space*. In that case, all functions from Ω to $\mathbb{R}$ will be random variables because you cannot have a pre-image which is not $\mathcal{F}$-measurable, So, only in cases where $\mathcal{F}$ is not 2^{Ω} is there a possibility that a function is not a random variable.

What we are interested now is - *what values X can take?* To be able to describe that, we will need to specify P_X for all Borel sets. Then, we will know the complete probabilistic description of the random variable on the real line. It turns out, you don't need P_X for all Borel sets. It is enough to specify P_X for certain *nice* sets. Let's see how we do that.

Recall when we were generating the Borel σ-algebra on the real line, we mentioned that there are multiple ways of generating it. One of them is using *open intervals*. The other important way is to use *semi-infinite intervals*. We can look at the $\mathcal{B}(\mathbb{R})$ as being generated by the sets of form $(-\infty, x]$ where x $\in \mathbb{R}$.

What are we trying to get at is - *the probability law that is defined for all Borel sets must also be defined for the generating class*. Thus, $P_X((-\infty, x])$ is well-defined for all $x \in \mathbb{R}$. Another way of writing $P_X((-\infty, x])$ is $P_X(\omega \in \Omega | X(\omega) \leq x)$. This probability has a special name called **cumulative distribution function (CDF)**, often denoted by $F_X(x)$.

So, if you provide the probability law for all Borel sets, we can easily figure out the CDF. What is not so obvious and requires a proof is *if you specify the CDF, you can specify the probability law for all Borel sets.* From the *Classical Probability*, we know that if we get the CDF of a random variable, we know everything about it. This is a correct concept but not so obvious. Because to know everything about the random variable, we must know probability measure of all the Borel sets. The reason this is possible is because all the other Borel sets you can write as the countable intersections and unions of the generating class, and we can figure out P_X from F_X. The way to formalize this notion is by using an object called *π-system.*

Definition 4.4. *Given a set Ω, a π-system on Ω is a non-empty collection of subsets of Ω that remains stable under finite intersections.*

Formally, if we denote this collection by $\mathcal{P}$, then $\mathcal{P}$ is a π-system if, for any subsets A and B in $\mathcal{P}$, the intersection $A \cap B$ also belongs to $\mathcal{P}$. This stability under intersection is a key property that defines π-systems.

Just like σ-algebra, π-system is another algebraic structure. π-systems require closure under only finite intersections - no complements or no countable unions. So, this structure is much weaker than σ-algebra. In fact, it is also much weaker than the *algebra*. All algebras are clearly π-systems, but not the other way around.

A particularly common example of a π-system on $\mathbb{R}$ (the set of real numbers) is the collection of all closed semi-infinite intervals of the form $(-\infty, x]$ where x is any real number. We denote this class as $\pi(\mathbb{R})$, and express it mathematically by:

$$\pi(\mathbb{R}) = \{(-\infty, x] : x \in \mathbb{R}\}.$$

This π-system is useful because each interval in $\pi(\mathbb{R})$ represents a closed set extending infinitely to the left and bounded by a specific real number x on the right. The intervals in $\pi(\mathbb{R})$ share the π-system property: if we take any two intervals $(-\infty, x_1]$ and $(-\infty, x_2]$ from $\pi(\mathbb{R})$, their intersection will also be of the same form, specifically $(-\infty, \min(x_1, x_2)]$, which again belongs to $\pi(\mathbb{R})$.

Lemma 4.1. *The **Borel** σ**-algebra** on a set R, often denoted by $\mathcal{B}(R)$, is the smallest σ-algebra containing all sets in a collection $\pi(R)$. We write this relation as:*

$$\mathcal{B}(R) = \sigma(\pi(R)),$$

*where $\sigma(\pi(R))$ represents the σ-algebra generated by $\pi(R)$. Here, $\pi(R)$ is called a π-**system** on R, which is simply a collection of subsets of R closed under finite intersections.*

To deepen our understanding, let us consider an important result from **measure theory** regarding such π-systems. Suppose we have two **finite measures**, say μ and ν, defined on a measurable space $(R, \sigma(\pi(R)))$. Now, if μ and ν **agree** on $\pi(R)$ (that is, $\mu(A) = \nu(A)$ for every $A \in \pi(R)$), then this result guarantees that μ and ν will also agree on the entire σ-algebra $\sigma(\pi(R))$. In other words, for every $B \in \sigma(\pi(R))$, it follows that:

$$\mu(B) = \nu(B).$$

This result is powerful because it tells us that to verify equality of two finite measures on the entire σ-algebra $\sigma(\pi(R))$, it suffices to verify their equality on the simpler structure of the π-system $\pi(R)$. This principle often simplifies work with measures by reducing the complexity of the verification needed, allowing us to focus on a smaller collection of sets.

Lemma 4.2. *Let Ω be a set and consider a π-system $\mathcal{P}$ over Ω. Specifically, if $A, B \in \mathcal{P}$, then $A \cap B \in \mathcal{P}$ as well. We denote by $\Sigma = \sigma(\mathcal{P})$ the σ-algebra generated by $\mathcal{P}$. The σ-algebra Σ includes all subsets of Ω that can be formed by countable unions, intersections, and complements of elements in $\mathcal{P}$.*

Proof. Let μ_1 and μ_2 be two measures defined on the measurable space (Ω, Σ) that agree on a few conditions:

1. Both measures are finite on Ω, meaning $\mu_1(\Omega) = \mu_2(\Omega) < \infty$.
2. The measures agree on $\mathcal{P}$, so $\mu_1(A) = \mu_2(A)$ for all $A \in \mathcal{P}$.

Our goal is to show that these conditions imply $\mu_1 = \mu_2$ on all of Σ, not just on $\mathcal{P}$.

The key idea here leverages the fact that $\mathcal{P}$ is a π-system. Intuitively, because $\mathcal{P}$ is closed under intersections, it forms a structure that allows the agreement of μ_1 and μ_2 on $\mathcal{P}$ to extend naturally to all of Σ. This result is a powerful extension theorem for measures, which can be formally stated as follows:

If two measures μ_1 and μ_2 agree on a π-system $\mathcal{P}$ and are finite on Ω, then they must agree on the entire σ-algebra $\Sigma = \sigma(\mathcal{P})$. That is,

$$\mu_1(E) = \mu_2(E) \quad \text{for all } E \in \Sigma.$$

Why does this work? The reasoning involves verifying that the set of all $E \in \Sigma$ for which $\mu_1(E) = \mu_2(E)$ forms a λ-system (also called a Dynkin system). A λ-system is a collection of sets closed under complements and countable disjoint unions, and every π-system is also contained within a λ-system that forms a σ-algebra. Thus, the combination of the π-system and the λ-system properties leads to the conclusion that $\mu_1 = \mu_2$ on Σ. □

Corollary 4.1. *If two probability measures agree on a π-system, then they agree on the σ-algebra generated by that π-system.*

We will use this result to prove the theorem on CDF.

4.1.1 Cumulative Distribution Function

Consider the *Probability Space* $(\Omega, \mathcal{F}, P)$ and a random variable $X : \Omega \to \mathbb{R}$. This mapping X induces a new probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ on the real line, where $\mathcal{B}(\mathbb{R}) = \sigma(\pi(\mathbb{R}))$ is the Borel σ-algebra generated by the collection of semi-infinite intervals on $\mathbb{R}$, i.e., the family $\{(-\infty, x] : x \in \mathbb{R}\}$.

For any $x \in \mathbb{R}$, we know:

$$(-\infty, x] \in \mathcal{B}(\mathbb{R}) \Rightarrow X^{-1}((-\infty, x]) \in \mathcal{F}.$$

Thus, it is meaningful to consider the probability measure P_X applied to these intervals, and we define the **Cumulative Distribution Function (CDF)** of X as:

$$F_X(x) = P_X((-\infty, x]) = P(\{\omega \in \Omega : X(\omega) \leq x\}), \quad x \in \mathbb{R}.$$

For convenience, we denote $P(X \leq x)$ as shorthand for $P(\{\omega \in \Omega : X(\omega) \leq x\})$, though this is slightly informal notation.

Theorem 4.2. *The probability law P_X of the random variable X is completely characterized by its CDF F_X. In other words, if two random variables X and Y have the same CDF, then they have the same probability law.*

Proof. Let $\mathcal{P} = \{(-\infty, x] : x \in \mathbb{R}\}$. This is a π-system on $\mathbb{R}$ because the intersection of any two intervals in $\mathcal{P}$ is again an interval in $\mathcal{P}$, i.e., if $x \leq y$, then:

$$(-\infty, x] \cap (-\infty, y] = (-\infty, \min(x, y)] \in \mathcal{P}.$$

The Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is generated by $\mathcal{P}$, so $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{P})$.

By the properties of a π-system and the extension theorem, any two probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that agree on the generating π-system $\mathcal{P}$ must agree on the entire σ-algebra $\mathcal{B}(\mathbb{R})$.

Since $F_X(x) = P_X((-\infty, x])$ uniquely determines the probability on each interval $(-\infty, x] \in \mathcal{P}$, the probability law P_X is fully determined by F_X. Therefore, the CDF F_X uniquely characterizes the probability distribution of the random variable X, as claimed. $\square$

4.1.2 Properties of CDF

Let X be a random variable with cumulative distribution function (CDF) $F_X(\cdot)$.

Theorem 4.3. *The CDF $F_X(x)$ is* ***monotonically non-decreasing****. Specifically, if $x \leq y$, then $F_X(x) \leq F_X(y)$.*

Proof. The CDF at a point x, $F_X(x)$, is the probability that the random variable X takes a value less than or equal to x. In other words,

$$F_X(x) = P(X \leq x).$$

Now, if $x \leq y$, we want to show that $F_X(x) \leq F_X(y)$.
This will follow from a property of sets and probability measures. Observe that the set of outcomes where $X \leq x$ is always contained within the set where $X \leq y$ whenever $x \leq y$. Formally, we have

$$\{\omega \mid X(\omega) \leq x\} \subseteq \{\omega \mid X(\omega) \leq y\}.$$

This inclusion tells us that every outcome that satisfies $X \leq x$ also satisfies $X \leq y$ when $x \leq y$.

Now, probability measures are **monotonic** with respect to sets: if one set is contained within another, then the probability of the smaller set is less than or equal to the probability of the larger set. Applying this to our situation, we get

$$P(X \leq x) \leq P(X \leq y).$$

Thus, we conclude that

$$F_X(x) \leq F_X(y),$$

proving that F_X is indeed a monotonically non-decreasing function. □

Theorem 4.4.

$$\lim_{x\to\infty} F_X(x) = 1 \quad \textit{and} \quad \lim_{x\to-\infty} F_X(x) = 0.$$

Proof. To understand this result, let's carefully examine what happens to the cumulative distribution function (CDF) $F_X(x) = P(X \leq x)$ as x approaches $-\infty$ and $+\infty$.

First, let's consider:

$$\lim_{x\to-\infty} F_X(x).$$

This limit represents the probability that X takes on a value less than or equal to a very large negative number, which we expect to approach zero. We can see why by breaking down the steps:

1. By definition of the CDF, we have:

$$\lim_{x\to-\infty} F_X(x) = \lim_{x\to-\infty} P(X \leq x).$$

2. Now, let's take a sequence $\{x_n\}_{n\in\mathbb{N}}$ where each x_n is smaller than the last, and this sequence decreases to $-\infty$. Thus:

$$\lim_{n\to\infty} P(X \leq x_n).$$

3. Observe that $P(X \leq x_n)$ corresponds to the event $\bigcap_{n\in\mathbb{N}}\{\omega : X(\omega) \leq x_n\}$, which becomes smaller and smaller as x_n goes to $-\infty$, eventually approaching the empty set, $\emptyset$. So, we get:

$$P\left(\bigcap_{n\in\mathbb{N}}\{\omega : X(\omega) \leq x_n\}\right) = P(\emptyset) = 0.$$

Thus:

$$\lim_{x\to-\infty} F_X(x) = 0.$$

Now, let's use similar reasoning to evaluate $\lim_{x\to\infty} F_X(x)$:

1. Again by the CDF definition,

$$\lim_{x\to\infty} F_X(x) = \lim_{x\to\infty} P(X \leq x).$$

2. Now consider a sequence $\{x_n\}_{n\in\mathbb{N}}$ where each x_n is greater than the last, and this sequence increases to $+\infty$:

$$\lim_{n\to\infty} P(X \leq x_n).$$

3. In this case, $P(X \leq x_n)$ corresponds to the event $\bigcup_{n\in\mathbb{N}}\{\omega : X(\omega) \leq x_n\}$, which approaches the entire sample space Ω as x_n becomes very large.

Thus:

$$P\left(\bigcup_{n\in\mathbb{N}}\{\omega : X(\omega) \le x_n\}\right) = P(\Omega) = 1.$$

Therefore:

$$\lim_{x\to\infty} F_X(x) = 1.$$

□

Theorem 4.5. *The cumulative distribution function $F_X(\cdot)$ is right-continuous, which means that for any $x \in \mathbb{R}$,*

$$\lim_{\varepsilon\downarrow 0} F_X(x+\varepsilon) = F_X(x).$$

Proof. To establish this, let's consider a sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ where each $\varepsilon_n > 0$ and $\varepsilon_n \to 0$ as $n \to \infty$. For any $x \in \mathbb{R}$, we proceed as follows:

$$\lim_{\varepsilon\downarrow 0} F_X(x+\varepsilon)$$

$$= \lim_{\varepsilon\downarrow 0} P(X \le x+\varepsilon) \quad \text{(from the definition of } F_X \text{ as the probability } P(X \le x)\text{)}.$$

Since $\varepsilon_n \to 0$, we can rewrite the limit using the sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$:

$$= \lim_{n\to\infty} P(X \le x+\varepsilon_n).$$

Now we use the continuity property of probability measures to take the limit over the decreasing sequence of events $\{\omega : X(\omega) \le x+\varepsilon_n\}$:

$$= P\left(\bigcap_{n\in\mathbb{N}}\{\omega : X(\omega) \le x+\varepsilon_n\}\right).$$

Since $\varepsilon_n \to 0$ implies $x+\varepsilon_n \to x$, the intersection above simplifies to the event $\{\omega : X(\omega) \le x\}$:

$$= P(X \le x) = F_X(x).$$

Thus, we have shown that

$$\lim_{\varepsilon\downarrow 0} F_X(x+\varepsilon) = F_X(x),$$

which completes the proof. □

Note: It's essential to note that while a cumulative distribution function (CDF) may not always be continuous, it must be **right-continuous** as shown above. This right-continuity is a fundamental property of all CDFs. Furthermore, any function that satisfies the properties of a CDF (non-decreasing, right-continuous, and approaching 0 as $x \to -\infty$ and 1 as $x \to \infty$) is guaranteed to be the CDF of some random variable. This result underscores the unique characteristics that any CDF must possess, tying it intrinsically to the behavior of random variables.

Theorem 4.6. *Let F be a function satisfying the three properties of a cumulative distribution function (CDF) as described in the last three theorems. Consider the probability space $\Omega = ([0,1), \mathcal{B}([0,1)), \lambda)$, where $\mathcal{B}([0,1))$ is the Borel sigma-algebra on the interval $[0,1)$ and λ is the Lebesgue measure. Then, there exists a random variable $X : \Omega \to \mathbb{R}$ whose CDF is F.*

Proof. We aim to construct a random variable X on Ω such that the CDF of X is precisely F. For this, let us first define the key properties of the function F and review the characteristics of our probability space Ω.

Since F is a CDF, it satisfies:

1. F is non-decreasing: for any $x_1, x_2 \in \mathbb{R}$ with $x_1 \leq x_2$, we have $F(x_1) \leq F(x_2)$.
2. F is right-continuous: for any $x \in \mathbb{R}$, $\lim_{y \to x^+} F(y) = F(x)$.
3. F has limits $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Given these properties, we can interpret $F(x)$ as the probability that a random variable X takes a value less than or equal to x.

To construct X explicitly, consider the probability space $\Omega = ([0,1), \mathcal{B}([0,1)), \lambda)$. For each $\omega \in \Omega$, we define $X(\omega)$ using the inverse function F^{-1}, where:

$$X(\omega) = F^{-1}(\omega) = \inf\{x \in \mathbb{R} : F(x) \geq \omega\}.$$

In this expression, $X(\omega)$ is defined as the smallest value x for which $F(x)$ is at least ω. This construction works because F is non-decreasing and right-continuous, ensuring that for each $\omega \in [0,1)$, such an x exists and is unique.

Now, let's verify that the CDF of X is indeed F. For any $x \in \mathbb{R}$, we want to compute the probability $\mathbb{P}(X \leq x)$. By the definition of X, we have:

$$\mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in \Omega : F^{-1}(\omega) \leq x\}).$$

Since $F^{-1}(\omega) \leq x$ if and only if $\omega \leq F(x)$, we can rewrite this as:

$$\mathbb{P}(X \leq x) = \mathbb{P}(\{\omega \in [0,1) : \omega \leq F(x)\}).$$

The probability of this event is simply the measure of the interval $[0, F(x))$ under λ, which is $F(x)$. Thus, we have shown:

$$\mathbb{P}(X \leq x) = F(x).$$

This confirms that F is indeed the CDF of X, as required. □

4.1.3 Indicator Random Variable

An **indicator random variable**, denoted as I_A, is a function that is defined on a probability space $(\Omega, \mathcal{F}, P)$. Specifically, for a measurable set $A \in \mathcal{F}$, the indicator random variable is defined as:

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \\ 0, & \text{if } \omega \notin A \end{cases}$$

This random variable takes the value 1 if the outcome ω falls within the set A, and 0 otherwise.

To establish that I_A is an f-measurable function, we need to show that for any $\lambda \in \mathbb{R}$, the set

$$\{\omega \in \Omega : I_A(\omega) \leq \lambda\}$$

is a measurable set in $\mathcal{F}$.

If $\lambda < 0$, then $\{\omega \in \Omega : I_A(\omega) \leq \lambda\} = \emptyset$, which is measurable.

If $0 \leq \lambda < 1$, then

$$\{\omega \in \Omega : I_A(\omega) \leq \lambda\} = \{\omega \in \Omega : I_A(\omega) = 0\} = \Omega \setminus A,$$

which is measurable since $A \in \mathcal{F}$.

If $\lambda \geq 1$, then

$$\{\omega \in \Omega : I_A(\omega) \leq \lambda\} = \Omega,$$

which is also measurable.

Thus, I_A is indeed f-measurable.

Next, let us consider the cumulative distribution function (CDF) of the indicator random variable I_A. The CDF, denoted as $F_{I_A}(x)$, is defined as follows:

$$F_{I_A}(x) = P(I_A \leq x).$$

We can evaluate this function in different scenarios:

If $x < 0$, then $F_{I_A}(x) = P(I_A \leq x) = 0$ because the indicator variable only takes values 0 or 1.
If $0 \leq x < 1$, then $F_{I_A}(x) = P(I_A = 0) = P(\Omega \setminus A) = 1 - P(A)$.
If $x \geq 1$, then $F_{I_A}(x) = P(I_A \leq x) = 1$ since I_A will always be less than or equal to x in this range.

Thus, we have the CDF of the indicator random variable I_A summarized as:

$$F_{I_A}(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - P(A), & \text{if } 0 \leq x < 1 \\ 1, & \text{if } x \geq 1 \end{cases}$$

4.2 Types of Random Variables

Random variables are classified based on the type of probability measure they induce on the real line. Formally, this measure is defined on the **Borel σ-algebra**, the collection of all sets that we consider *measurable* on the real line. Now, it turns out that there are three fundamentally distinct ways a probability measure can behave on the real line. An important result in measure theory, known as the **Lebesgue Decomposition Theorem** (we will see this theorem later), guarantees that any probability measure on $\mathbb{R}$ can be uniquely expressed as a combination of these three types of measures. These foundational types are Discrete, Continuous and Singular. Thus, we say that there are three *pure type* random variables, namely: **discrete random variables**, **continuous random variables**, and **singular random variables**. Furthermore, it is possible to construct random variables by combining these pure types, resulting in what we call **mixed random variables**. In total, this yields seven possible types of random variables, including these mixtures.

However, for most applications—especially in engineering and statistics—only **discrete** and **continuous random variables** are practically significant. **Singular random variables** are rarely encountered outside of theoretical discussions and are mostly studied in academia. Consequently, our primary focus will be on understanding discrete and continuous random variables, though we will also define and illustrate an example of a singular random variable for completeness.

4.2.1 Discrete Random Variables

Definition 4.5. *A random variable X is called discrete if it takes values in a countable subset of $\mathbb{R}$ with probability 1. In other words, there exists a countable set $E = \{x_1, x_2, \dots\}$ such that $P_X(E) = 1$, meaning the entire probability distribution of X is contained within this set.*

To clarify, we don't necessarily require that the *range* of X itself is countable; only that X almost surely takes values in a countable subset. There may still be some subset of the sample space that maps to an uncountable subset of $\mathbb{R}$ but with probability zero.

Definition 4.6. *Given a discrete random variable X, we define the function $p_X : \mathbb{R} \to [0,1]$ by*

$$p_X(x) = P(X = x) \quad \textit{for every } x.$$

This function p_X is called the probability mass function (PMF) of X. While $p_X(x)$ is defined for all $x \in \mathbb{R}$, it's important to observe that $p_X(x)$ is only non-zero for $x \in E$, the countable set where X has probability mass.

Since $P_X(E) = 1$, the additivity property of probabilities implies:

$$\sum_{i=1}^{\infty} P(X = x_i) = 1.$$

This summation states that the total probability across all possible values X can take must equal 1.

For a discrete random variable X, the PMF p_X alone provides a complete description of the probability distribution of X. For any Borel set $B \subset \mathbb{R}$, we can determine the probability that X falls within B by summing over the values in E that are also in B:

$$P_X(B) = \sum_{i : x_i \in B} P(X = x_i).$$

This relation highlights that for any subset B of the real numbers, we only need to consider the countable values where the PMF is non-zero.

For a discrete random variable, the cumulative distribution function (CDF) F_X at any point x is given by the sum of probabilities for all values $x_i \leq x$:

$$F_X(x) = \sum_{i : x_i \leq x} P(X = x_i).$$

This CDF describes the probability that X takes a value less than or equal to x, accumulated over all points in E up to x.

To illustrate some of the most common discrete random variables, we delve into a few key examples below, each defined on a probability space $(\Omega, \mathcal{F}, P)$.

1. Indicator Random Variable:

Let $A \in \mathcal{F}$ be any event. We define the indicator random variable $I_A : \Omega \to \{0, 1\}$ by

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases}$$

The variable I_A is indeed a random variable because both A and A^c are $\mathcal{F}$-measurable. As it only takes two values (0 and 1), it is clearly discrete.

2. Bernoulli Random Variable:

Consider a parameter $p \in [0,1]$. We define a random variable X with probability mass function (PMF) given by

$$P(X=0) = p \quad \text{and} \quad P(X=1) = 1-p.$$

This random variable can represent a single coin toss: $X = 0$ can signify *heads* with probability p, and $X = 1$ can represent *tails* with probability $1-p$. If $p = \frac{1}{2}$, the coin is fair.

3. Discrete Uniform Random Variable:

Suppose a and b are integers with $a < b$. Define a random variable X with values $m = a, a+1, \ldots, b$ and probability

$$P(X=m) = \frac{1}{b-a+1}.$$

This variable takes each integer between a and b inclusively with equal probability. Outside this range, $P(X=m) = 0$.

4. Binomial Random Variable:

Given a parameter $p \in [0,1]$ and an integer $n \in \mathbb{N}$, a binomial random variable X counts the number of successes in n independent trials, each with success probability p. Its PMF is

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n.$$

This model applies to scenarios like counting the number of heads in n independent coin tosses with success probability p.

5. Geometric Random Variable:

For $0 < p \leq 1$, the geometric random variable X represents the number of independent trials required to obtain the first success. Its PMF is given by

$$P(X=k) = p(1-p)^{k-1} \quad \text{for } k = 1, 2, \ldots$$

This model could describe, for instance, the number of coin tosses until the first head appears, where each toss has a success probability of p.

6. Poisson Random Variable:

Given a parameter $\lambda > 0$, a Poisson random variable X represents the number of occurrences of an event in a fixed interval, assuming events occur with a constant mean rate λ. The PMF of X is

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

The Poisson model is often used for count data, such as the number of emails received per hour.

Except for the indicator random variable, each of these examples is primarily defined by its PMF rather than an explicit mapping from the sample space Ω.

4.2.2 Continuous Random Variables

To properly define continuous random variables, we need to understand the concept of **absolute continuity** between measures.

Definition 4.7. *Let μ and ν be measures on $(\Omega, \mathcal{F})$. We say that ν is absolutely continuous with respect to μ if for every set $N \in \mathcal{F}$ where $\mu(N) = 0$, we also have $\nu(N) = 0$. In other words, if μ assigns no weight or size to N, then ν must do the same.*

Now, consider a probability space $(\Omega, \mathcal{F}, P)$ and let $X : \Omega \to \mathbb{R}$ represent a random variable. We call X a **continuous random variable** if the *law* (or distribution) of X, denoted P_X, is absolutely continuous with respect to the *Lebesgue measure* λ on $\mathbb{R}$.

Let's clarify the terms:

(a) The law P_X is a measure on the real line $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the collection of Borel sets on $\mathbb{R}$.
(b) Absolute continuity of P_X with respect to λ means that for any Borel set $N \subseteq \mathbb{R}$ with $\lambda(N) = 0$, we must also have $P_X(N) = P(\omega \mid X(\omega) \in N) = 0$.

This ensures that a continuous random variable X does not place probability mass on sets of *Lebesgue measure zero,* such as individual points or countable collections of points.

Importantly, just because a random variable X takes values in an uncountable set does not mean it is continuous. Continuity here is defined in terms of the behavior of the probability measure P_X in relation to the Lebesgue measure λ.

To formalize this relationship, we rely on a powerful tool called the **Radon-Nikodym Theorem**. Though we won't delve into the proof right now, this theorem asserts that:

Theorem 4.7. *If P_X is absolutely continuous with respect to λ, there exists a non-negative, measurable function f_X, known as the Radon-Nikodym derivative of P_X with respect to λ, such that:*

$$P_X(A) = \int_A f_X(x)\, d\lambda(x)$$

for any Borel set $A \subseteq \mathcal{B}(\mathbb{R})$.

This function f_X essentially describes the *density* of the probability measure P_X relative to λ, giving us a precise way to calculate probabilities over continuous random variables.

The integral used in the theorem above differs from the typical Riemann integral. This is because the set B can be any Borel measurable set, such as the Cantor set, not necessarily an interval. In this course, we will later dive into abstract integration to achieve a precise understanding of the integral. For now, however, we can consider B as a simple interval $[a, b]$.

Under this assumption, equation tells us that the probability of the random variable X falling within the interval $[a, b]$ can be expressed as:

$$\int_a^b f_X\, dx$$

where f_X is a non-negative measurable function.

When we say that f_X is measurable, we mean specifically that if we take the pre-images of Borel sets under f_X, these pre-images are also Borel sets. This property ensures that f_X is compatible with the Borel structure, making the integration meaningful over any Borel set B.

Below, we discuss some prominent examples of continuous random variables, each with unique properties and interpretations.

1. Uniform Distribution:

The uniform distribution on a closed interval $[a, b]$ is one where each point in the interval has an equal probability density. It is often described as a *scaled Lebesgue measure* over $[a, b]$.

(a) *Probability Density Function (PDF)*:

$$f_X(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{for } x > b \end{cases}$$

This function shows that outside the interval $[a, b]$, the probability density is zero, and within $[a, b]$, the density is constant.

(b) *Cumulative Distribution Function (CDF)*:

$$F_X(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \leq x \leq b \\ 1 & \text{for } x > b \end{cases}$$

The CDF $F_X(x)$ gives the probability that X is less than or equal to x, increasing linearly within the interval and reaching 1 at $x = b$.

2. Exponential Distribution:

The exponential distribution is used to model the time until an event occurs and is defined for $x \geq 0$, characterized by a parameter $\lambda > 0$. It is unique in that it has a special property known as the **memoryless property**.

(a) *Probability Density Function (PDF)*:

$$f_X(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

Here, λ determines the rate at which probabilities decay as x increases.

(b) *Cumulative Distribution Function (CDF)*:

$$F_X(x) = 1 - e^{-\lambda x} \quad \text{for } x \geq 0$$

This CDF gives the probability that X is less than or equal to x and approaches 1 as x grows, meaning that the event is more likely to have occurred by larger values of x.

Definition 4.8. *Memoryless Property states that the probability of an event occurring in the future is independent of the time that has already elapsed. Formally, a non-negative random variable X is said to be memoryless if:*

$$P(X > s + t \mid X > t) = P(X > s) \quad \forall s, t \geq 0.$$

To show that an exponential random variable satisfies this property, consider:

$$P(X > s + t \mid X > t) = \frac{P((X > s + t) \cap (X > t))}{P(X > t)} = \frac{P(X > s + t)}{P(X > t)}.$$

Substituting the exponential form, we get:

$$\frac{e^{-(s+t)\lambda}}{e^{-t\lambda}} = e^{-s\lambda} = P(X > s).$$

Thus, the exponential distribution indeed possesses the memoryless property.

This property has a practical interpretation. For example, if the lifetime of a light bulb is exponentially distributed, then the expected time to failure, given that the bulb has not yet failed by time t, is the same as the expected lifetime of a completely new bulb. Remarkably, the exponential distribution is the only continuous distribution that exhibits this property.

3. Normal/Gaussian Distribution:

The Gaussian (or Normal) distribution is a fundamental two-parameter distribution, where the parameters are:

(a) The mean $\mu \in \mathbb{R}$, representing the center of the distribution.
(b) The standard deviation $\sigma > 0$, which reflects the spread or dispersion of the distribution.

Gaussian distributions are ubiquitous in engineering, statistics, and the natural sciences, largely because they possess a *stable-attractor* property.

Definition 4.9. *Stable attractor property for Gaussian distributions means that sums of independent Gaussian random variables tend to result in a Gaussian distribution.*

This is a major reason for their prominence. We will explore this property and its implications in more detail later.
(a) *Probability Density Function (PDF)*:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

This distribution is commonly denoted as $N(\mu, \sigma^2)$, where μ is the mean and σ^2 is the variance (the square of the standard deviation). A particularly notable case occurs when $\mu = 0$ and $\sigma^2 = 1$; this is called the *standard Gaussian distribution*, and its PDF simplifies to:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

(b) *Cumulative Distribution Function (CDF)*: The cumulative distribution function (CDF) of a Gaussian random variable does not have a closed-form expression, which might seem inconvenient. However, this lack of a *closed-form* solution is often manageable and does not detract from its usefulness. For convenience, the CDF of the standard Gaussian (where $\mu = 0$ and $\sigma = 1$) is given a special name: the *error function*, denoted as $\mathrm{Erf}(x)$, and defined by:

$$\mathrm{Erf}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy.$$

This integral essentially captures the probability that a standard Gaussian random variable falls within a particular range, from $-\infty$ to x, and is widely used in statistics and probability theory.

4. Cauchy Distribution:

The Cauchy distribution, characterized by two parameters $x_0 \in \mathbb{R}$ (the location parameter) and $\gamma > 0$ (the scale parameter), is a notable example of a **heavy-tailed** distribution.

Definition 4.10. *Heavy-tailed nature means that the probability of extreme values does not diminish as rapidly as it would in distributions with lighter tails, like the normal distribution.*

As a result, the Cauchy distribution is commonly used in engineering and other fields to model phenomena with high variability or *burstiness.*

(a) *Probability Density Function (PDF)*:

$$f_X(x) = \frac{1}{\pi} \frac{\gamma}{(x - x_0)^2 + \gamma^2}.$$

(b) *Cumulative Distribution Function (CDF)*: To find the cumulative distribution function (CDF) $F_X(x)$, which represents the probability that $X \leq x$, we integrate the PDF from $-\infty$ to x:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{x} \frac{1}{\pi} \frac{\gamma}{(t - x_0)^2 + \gamma^2}\, dt.$$

This integral yields:

$$F_X(x) = \frac{1}{\pi} \arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}.$$

Thus, the cumulative distribution function of the Cauchy distribution is:

$$F_X(x) = \frac{1}{\pi} \arctan\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}.$$

This CDF is defined over the entire real line and illustrates the heavy-tailed nature of the Cauchy distribution. Notice that as $x \to \infty$, $F_X(x) \to 1$, while as $x \to -\infty$, $F_X(x) \to 0$, consistent with the fact that this is a probability distribution over $\mathbb{R}$.

4.2.3 Singular Random Variables

Singular random variables are fascinating and unusual objects in probability theory, often seeming to defy intuition. They are distinct because they exist in a realm that neither discrete nor continuous random variables fully describe. Specifically, singular random variables take on values with probability one within an uncountable set that has a Lebesgue measure of zero.

Definition 4.11. *A random variable X is **singular** if, for any $x \in \mathbb{R}$, we have $P_X(\{x\}) = 0$, and yet there exists a set $F \in \mathcal{B}(\mathbb{R})$ (where $\mathcal{B}(\mathbb{R})$ denotes the Borel sets on $\mathbb{R}$) such that F has zero Lebesgue measure and $P_X(F) = 1$.*

Why must F be uncountable?

Suppose F were countable; then it would only contain a finite or countably infinite number of points, meaning that any random variable with values in F would be akin to a discrete variable, contradicting our definition of $P_X(\{x\}) = 0$ for each $x \in F$. Therefore, F must be uncountable, though it still occupies *no space* in terms of Lebesgue measure.

Example: The Cantor Distribution as a Singular Random Variable

An example of a singular random variable arises with the Cantor distribution, whose cumulative distribution function (CDF) is known as the *Cantor function* or sometimes the *Devil's staircase*. The values of this random variable are taken from the Cantor set C, which is uncountable but has Lebesgue measure zero.

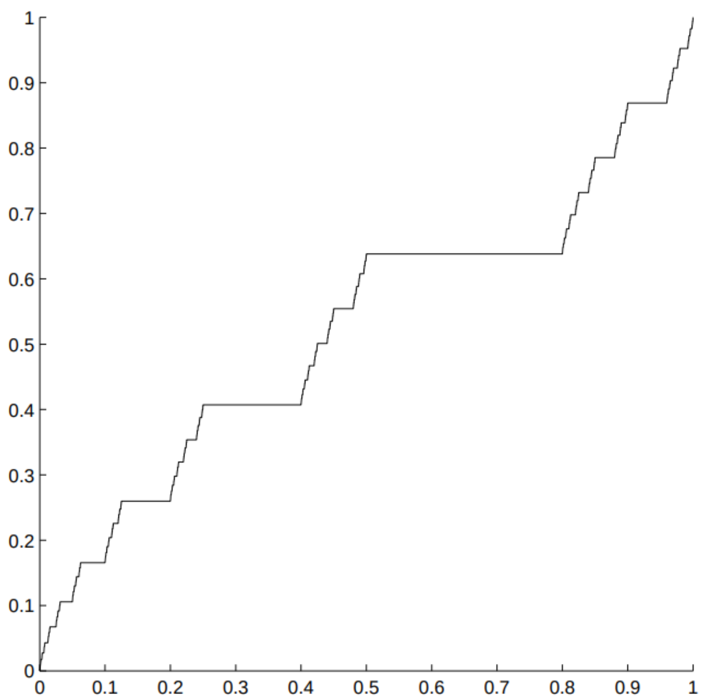
The Cantor set C is constructed by repeatedly removing the middle third from each interval in $[0, 1]$. Any $x \in C$ has a unique representation in a ternary (base-3) expansion, where each digit x_i can only be 0 or 2, such that:

$$x = \sum_{i=1}^{\infty} \frac{x_i}{3^i}, \quad x_i \in \{0, 2\}.$$

To construct a random variable X with values in the Cantor set, consider an infinite sequence of fair coin tosses. Assign $x_i = 2$ if the i-th toss is a head and $x_i = 0$ if it is a tail. Using the values x_i, we form a number x according to the series above. This defines a random variable X that takes values in the Cantor set C.

This X fulfills the two criteria for being a singular random variable:

1. $P_X(C) = 1$, meaning that X takes values entirely within the Cantor set.
2. $P_X(\{x\}) = 0$ for any specific $x \in C$, as each point in the Cantor set has zero probability.



Properties of the Cantor Function (CDF of X)

(a) It is **continuous everywhere** across $[0, 1]$, meaning it has no jump discontinuities.
(b) Its derivative is **zero almost everywhere**, reflecting that the function has a flat slope at almost every point in $[0, 1]$ except on the Cantor set itself.

(c) At points in the Cantor set C, the function increases in value, yet it does so in a manner without a well-defined derivative at those points—illustrating its *staircase* nature without abrupt jumps.

The Cantor function thus *climbs* incrementally across the interval $[0, 1]$ by increasing only at points in the Cantor set. This results in a singular CDF that is entirely smooth yet almost everywhere flat, a distinctive feature setting it apart from typical continuous or discrete distributions.

Exercise 4.1. *Consider a random variable X. Prove that*

$$P_X(\{y\}) = F_X(y) - \lim_{x \uparrow y} F_X(x).$$

Furthermore, show that the cumulative distribution function F_X is continuous at the point y if and only if the probability that X takes the value y, denoted as $P_X(\{y\})$, is equal to zero.

Solution 4.1. *Let X be a random variable, and we want to prove the following statement:*

$$P_X(\{y\}) = F_X(y) - \lim_{x \uparrow y} F_X(x)$$

where $F_X(y)$ denotes the cumulative distribution function (CDF) of X, which is defined as

$$F_X(y) = P_X(X \leq y).$$

To understand the left-hand side, $P_X(\{y\})$, we interpret it as the probability that the random variable X takes on the specific value y. This can be thought of as the measure of the set $\{y\}$ under the probability measure defined by X.

Now, consider the right-hand side of the equation. The term $F_X(y)$ represents the probability that X is less than or equal to y. In contrast, $\lim_{x \uparrow y} F_X(x)$ gives us the probability that X is less than y as we approach y from the left.

Thus, we can interpret the difference $F_X(y) - \lim_{x \uparrow y} F_X(x)$ as the amount of probability mass located precisely at y. If there is a non-zero probability that X equals y, then the CDF will increase at that point, resulting in a positive difference.

Next, we will show that F_X is continuous at y if and only if $P_X(\{y\}) = 0$.

(a) If F_X is continuous at y:

$$\text{Then } \lim_{x \uparrow y} F_X(x) = F_X(y).$$

Therefore,

$$P_X(\{y\}) = F_X(y) - \lim_{x \uparrow y} F_X(x) = F_X(y) - F_X(y) = 0.$$

(b) Conversely, if $P_X(\{y\}) = 0$: This implies that

$$F_X(y) - \lim_{x \uparrow y} F_X(x) = 0.$$

Hence,

$$F_X(y) = \lim_{x \uparrow y} F_X(x).$$

This shows that F_X is continuous at y.

In summary, we have established that $P_X(\{y\}) = F_X(y) - \lim_{x \uparrow y} F_X(x)$, and that F_X is continuous at y if and only if $P_X(\{y\}) = 0$.

Exercise 4.2. *Among the functions given below, identify which are valid cumulative distribution functions (CDFs) and find their corresponding densities. For those that are not valid CDFs, explain the failures.*
(a)

$$F(x) = \begin{cases} 1 - e^{-x^2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(b)

$$F(x) = \begin{cases} e^{-\frac{1}{x}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

(c)

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{3} & 0 < x \leq 1 \\ 2 & x > 1 \end{cases}$$

Solution 4.2. ***(a) Analysis:***

(a) For $x < 0$, $F(x) = 0$.
(b) For $x \geq 0$, as $x \to 0$, $F(0) = 1 - e^0 = 0$, and as $x \to \infty$, $F(x) \to 1$.
(c) The function is non-decreasing for $x \geq 0$ since the derivative $F'(x) = 2xe^{-x^2}$ is non-negative.

Thus, $F(x)$ is a valid CDF.

Finding the Density: *The PDF is given by the derivative:*

$$f(x) = \frac{d}{dx}F(x) = \begin{cases} 2xe^{-x^2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(b) Analysis:

(a) For $x \leq 0$, $F(x) = 0$.
(b) As $x \to 0^+$, $F(x) \to 0$, but as $x \to \infty$, $F(x) \to 1$.
(c) However, the function is non-decreasing for $x \geq 0$ because the derivative $F'(x) = e^{-\frac{1}{x}} \cdot \frac{1}{x^2}$ is always positive.

Thus, $F(x)$ is a valid CDF.

Finding the Density: *The PDF is given by the derivative:*

$$f(x) = \frac{d}{dx}F(x) = \begin{cases} e^{-\frac{1}{x}} \cdot \frac{1}{x^2} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

(c) Analysis:

(a) For $x \leq 0$, $F(x) = 0$.
(b) For $0 < x \leq 1$, $F(x) = \frac{1}{3}$, which is constant and thus non-decreasing.
(c) However, for $x > 1$, $F(x) = 2$, which violates the upper limit condition of a CDF, as $F(x)$ must approach 1 as $x \to \infty$.

Thus, $F(x)$ is not a valid CDF due to exceeding the limit of 1.

Exercise 4.3. ***Negative Binomial Random Variable.*** *Consider a sequence of independent Bernoulli trials $\{X_i\}_{i\in\mathbb{N}}$ with parameter of success $p \in (0,1]$. The number of successes in the first n trials is given by*

$$Y_n = \sum_{i=1}^{n} X_i.$$

Y_n is distributed as Binomial with parameters n and p. Consider the random variable defined by

$$V_k = \min\{n \in \mathbb{N}^+ : Y_n = k\}.$$

Note that V_1 is distributed as Geometric with parameter p.

(a) Give a verbal description of the random variable V_k

(b) Show that the probability mass function of the random variable V_k is given by

$$P(V_k = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k},$$

where $n \in \{k, k+1, \dots\}$, we need to consider the conditions under which $V_k = n$.

(c) Argue that the Binomial and Negative Binomial Distributions are inverse to each other in the sense that

$$Y_n \geq k \Leftrightarrow V_k \leq n,$$

Solution 4.3. ***(a)*** *A verbal description of the random variable V_k is as follows: V_k represents the number of trials required to achieve exactly k successes in a series of independent Bernoulli trials, where each trial has a success probability of p. In simpler terms, V_k tells us how many attempts we need to make until we obtain k successful outcomes.*

(b) *For V_k to equal n, the following must hold:*

1. The n-th trial must result in a success, contributing to the k-th success.
2. Among the first $n-1$ trials, exactly $k-1$ must be successes, ensuring that the k-th success occurs on the n-th trial.

The number of ways to select which $k-1$ trials out of the first $n-1$ are successful is given by $\binom{n-1}{k-1}$. The probability of exactly $k-1$ successes and $n-k$ failures in the first $n-1$ trials is given by $p^{k-1}(1-p)^{(n-1)-(k-1)}$ or $p^{k-1}(1-p)^{n-k}$. Finally, the n-th trial must be a success, contributing a factor of p. Therefore, the total probability is:

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

This is known as the Negative Binomial Distribution with parameters k and p.

(c) *To argue that the Binomial and Negative Binomial Distributions are inverse to each other in the sense that*

$$Y_n \geq k \Leftrightarrow V_k \leq n,$$

we interpret these statements as follows:

(a) $Y_n \geq k$ means that in n trials, we have at least k successes. This implies that the k-th success can occur at or before the n-th trial.

(b) $V_k \leq n$ means that the k-th success occurs within the first n trials. This is equivalent to stating that we need n or fewer trials to achieve k successes.

Thus, the two statements describe the same event: having k successes in at most n trials. Consequently, we can conclude that $Y_n \geq k \Leftrightarrow V_k \leq n$.

Exercise 4.4. *Radioactive decay. Assume that a radioactive sample emits a random number of α particles in any given hour, and that the number of α particles emitted in an hour is Poisson distributed with parameter λ. Suppose that a faulty Geiger-Muller counter is used to count these particle emissions. In particular, the faulty counter fails to register an emission with probability p, independently of other emissions.*

(a) What is the probability that the faulty counter will register exactly k emissions in an hour?
(b) Given that the faulty counter registered k emissions in an hour, what is the PMF of the actual number of emissions that happened from the source during that hour?

Solution 4.4. *To solve this problem, we will consider the nature of radioactive decay and the behavior of the faulty counter.*

First, we know that the number of α particles emitted in an hour, denoted as X, follows a Poisson distribution with parameter λ:

$$P(X = n) = \frac{\lambda^n e^{-\lambda}}{n!}$$

for $n = 0, 1, 2, \ldots$.

The faulty counter registers an emission with probability $1 - p$ and fails to register it with probability p. Therefore, if X is the actual number of emissions, the number of emissions registered by the counter, denoted as Y, follows a Binomial distribution:

$$Y | X = n \sim Binomial(n, 1 - p).$$

***(a)** We want to find the probability that the faulty counter registers exactly k emissions in an hour, which can be expressed as:*

$$P(Y = k) = \sum_{n=k}^{\infty} P(Y = k | X = n) P(X = n).$$

The conditional probability $P(Y = k | X = n)$ for the Binomial distribution is given by:

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

Thus, we have:

$$P(Y = k) = \sum_{n=k}^{\infty} \binom{n}{k} (1 - p)^k p^{n-k} \frac{\lambda^n e^{-\lambda}}{n!}.$$

This can be simplified by recognizing that:

$$\binom{n}{k} \frac{1}{n!} = \frac{1}{k!(n - k)!}.$$

The expression for $P(Y = k)$ becomes:

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

The inner sum is the series expansion for the exponential function, leading to:

$$\sum_{m=0}^{\infty} \frac{(\lambda p)^m}{m!} = e^{\lambda p}.$$

Therefore, we obtain:

$$P(Y = k) = (1-p)^k \frac{(\lambda(1-p))^k e^{-\lambda}}{k!}.$$

This shows that Y also follows a Poisson distribution with parameter $\lambda(1-p)$:

$$P(Y = k) = \frac{(\lambda(1-p))^k e^{-\lambda(1-p)}}{k!}.$$

(b) *Next, we need to find the PMF of the actual number of emissions X given that the counter registered k emissions:*

$$P(X = n|Y = k).$$

Using Bayes' theorem, we can express this as:

$$P(X = n|Y = k) = \frac{P(Y = k|X = n)P(X = n)}{P(Y = k)}.$$

Substituting the earlier expressions, we find:

$$P(X = n|Y = k) = \frac{\binom{n}{k}(1-p)^k p^{n-k} \frac{\lambda^n e^{-\lambda}}{n!}}{P(Y = k)}.$$

Given that $P(Y = k)$ has already been derived, we can substitute this back into our equation. This results in a formula that allows us to compute the conditional probabilities depending on the values of n and k.

In summary, the solution outlines the Poisson nature of emissions and the impact of a faulty counter on the observed counts.

Exercise 4.5. *Buses arrive at ten minute intervals starting at noon. A man arrives at the bus stop at a random time X minutes after noon, where X has the CDF:*

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{60} & 0 \leq x \leq 60 \\ 1 & x > 60. \end{cases}$$

What is the probability that he waits less than five minutes for a bus?

Solution 4.5. *To solve this problem, we first need to understand the situation described. The man arrives at the bus stop at a random time X uniformly distributed between 0 and 60 minutes after noon. The buses arrive every 10 minutes, which means they arrive at 0, 10, 20, 30, 40, 50, and 60 minutes. The key is to find the probability that the man waits less than 5 minutes for the next bus.*

The waiting time W can be calculated based on when he arrives at the bus stop:

1. *If $X \in [0, 5)$, the wait time $W < 5$.*
2. *If $X \in [5, 10)$, then $W = 10 - X < 5$ corresponds to $X > 5$, which means he waits for 0 to 5 minutes.*
3. *This pattern continues for $X \in [10, 15)$, where he will again have a wait time of less than 5 minutes.*

More formally, we can analyze the intervals for which $W < 5$: X in the intervals $[0,5)$ and $[10,15)$, and so on up to $[50,55)$.

To summarize the intervals where $W < 5$:

$$X \in [0,5) \cup [10,15) \cup [20,25) \cup [30,35) \cup [40,45) \cup [50,55).$$

Thus, the total length of intervals where the man waits less than 5 minutes is:

$$6 \times 5 = 30 \textit{ minutes}.$$

Since X is uniformly distributed over the interval $[0,60]$, the probability $P(W < 5)$ is the ratio of the total length of intervals where he waits less than 5 minutes to the total possible time (60 minutes):

$$P(W < 5) = \frac{30}{60} = \frac{1}{2}.$$

Exercise 4.6. *Find the values of a and b such that the following function is a valid CDF:*

$$F(x) = \begin{cases} 1 - ae^{-x/b} & \textit{for } x \geq 0 \\ 0 & \textit{for } x < 0 \end{cases}$$

Also, find the values of a and b such that the function above corresponds to the CDF of some:

1. *Continuous Random Variable*
2. *Discrete Random Variable*
3. *Mixed type Random Variable*

Solution 4.6. *To determine the values of a and b that make $F(x)$ a valid cumulative distribution function (CDF), we need to ensure that:*

1. $F(x)$ is non-decreasing.
2. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

For the given function $F(x)$:

1. For $x < 0$, we have $F(x) = 0$, which satisfies the first condition.
2. For $x \geq 0$, we have $F(x) = 1 - ae^{-x/b}$.

Next, we examine the limits:

(a) As $x \to 0$:

$$F(0) = 1 - ae^0 = 1 - a$$

To ensure that $F(0)$ is non-negative, we require:

$$1 - a \geq 0 \implies a \leq 1$$

(b) As $x \to \infty$:

$$F(x) \to 1 - a \cdot 0 = 1$$

This is valid if $a > 0$ to ensure $F(x)$ approaches 1 correctly.

Thus, from these conditions, we have:

$$0 < a \leq 1$$

Next, for b, the exponential function $e^{-x/b}$ is defined for all x when $b > 0$. Therefore, we must have:

$$b > 0$$

In summary, the parameters a and b must satisfy:

$$0 < a \leq 1 b > 0$$

Now, let's analyze the types of random variables.

(a) Continuous Random Variable

For $F(x)$ to represent the CDF of a continuous random variable, the function $F(x)$ must be strictly increasing. This requires:

$$a > 0 \quad \text{and} \quad b > 0$$

Therefore, the same conditions hold.

(b) Discrete Random Variable

For a discrete random variable, $F(x)$ should have jumps at specific values. In this case, we can choose $a = 1$ and b can be any positive number. Thus:

$$a = 1, \quad b > 0$$

(c) Mixed Random Variable

For a mixed type random variable, $F(x)$ must be a combination of both continuous and discrete. We can choose $a < 1$ and $b > 0$. This allows $F(x)$ to have a continuous component as well as a discrete one at $x = 0$. Thus, we can set:

$$0 < a < 1, \quad b > 0$$

Exercise 4.7. *Let X be a continuous random variable. Show that X is memoryless if and only if X is an exponential random variable.*

Solution 4.7. *To demonstrate that a continuous random variable X is memoryless if and only if it is an exponential random variable, we first define what it means for X to be memoryless. A random variable X is said to be memoryless if it satisfies the following condition for all $s, t \geq 0$:*

$$P(X > s + t \mid X > s) = P(X > t).$$

This equation states that the probability that X exceeds $s + t$, given that it has already exceeded s, is the same as the probability that X exceeds t alone.

(⇒) We begin by assuming that X is memoryless. To show that X is an exponential random variable, we will derive its cumulative distribution function (CDF) and probability density function (PDF).

Let $F(x) = P(X \leq x)$ be the CDF of X. Consequently, the survival function, which gives the probability that X exceeds x, is:

$$S(x) = P(X > x) = 1 - F(x).$$

Now, using the memoryless property, we have:

$$P(X > s + t \mid X > s) = \frac{P(X > s + t)}{P(X > s)}.$$

Using the memoryless condition, we equate this to $P(X > t)$:

$$\frac{S(s+t)}{S(s)} = S(t).$$

Rearranging gives us:

$$S(s+t) = S(s)S(t).$$

This functional equation resembles the form of the survival function of an exponential distribution. If we denote $S(t) = e^{-\lambda t}$ for some $\lambda > 0$, then we can verify that:

$$S(s+t) = e^{-\lambda(s+t)} = e^{-\lambda s}e^{-\lambda t} = S(s)S(t).$$

Thus, $S(t)$ takes the form of the exponential survival function, confirming that X is indeed an exponential random variable.

($\Leftarrow$) Conversely, suppose that X is an exponential random variable with parameter λ. The survival function is given by:

$$S(t) = P(X > t) = e^{-\lambda t}.$$

We can apply this to check the memoryless property:

$$P(X > s+t \mid X > s) = \frac{P(X > s+t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t).$$

Since this holds for all $s, t \geq 0$, we conclude that the exponential distribution satisfies the memoryless property.

In summary, we have shown that X is memoryless if and only if X is an exponential random variable.

4.3 Multiple Random Variables

We begin by examining multiple random variables defined on a common probability space. Let's focus on two random variables, X and Y, which are defined on the probability space $(\Omega, \mathcal{F}, P)$. It's crucial to grasp that the values taken by X and Y are influenced by the same underlying randomness, represented by $\omega \in \Omega$.

For instance, consider a scenario involving weather on a specific day. We can let the random variable X represent the temperature of that day, while another random variable Y denotes the humidity level. Since both X and Y are determined by the same outcome, it is natural to expect a certain level of interdependence between them. In this weather example, a higher temperature generally correlates with increased humidity.

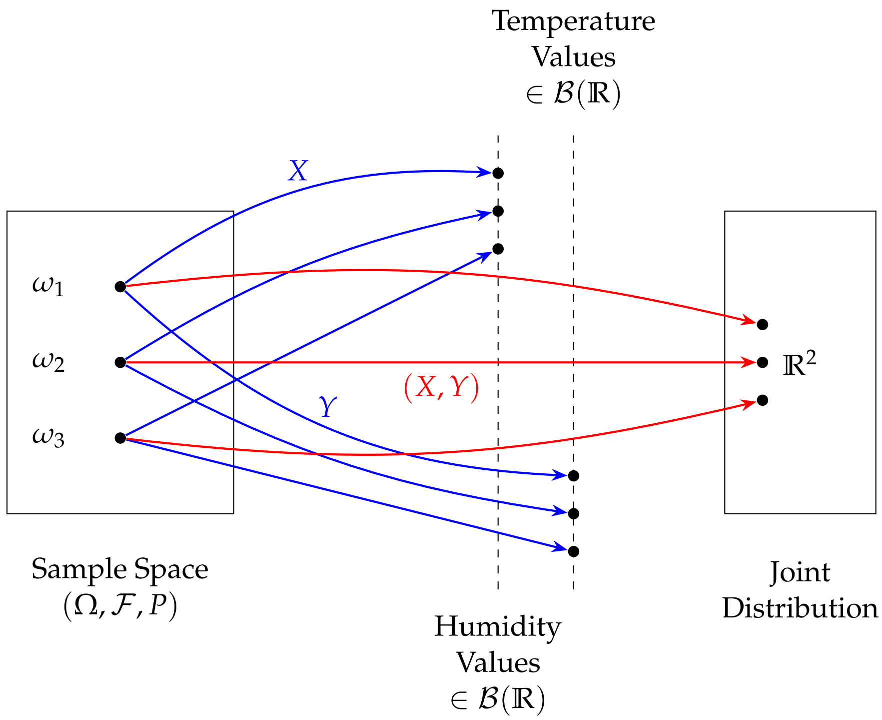
Say X and Y serve as measurable functions from the same probability space to the real numbers. The figure provided represents the pair $(X(\cdot), Y(\cdot))$ as a function mapping Ω to $\mathbb{R}^2$. This representation is particularly significant as it captures the interdependence between X and Y.

A pertinent question arises: is the function $(X(\cdot), Y(\cdot)) : \Omega \to \mathbb{R}^2$ measurable, given that both X and Y are measurable functions? To properly address this question, we first need to define the Borel σ-algebra on $\mathbb{R}^2$. The Borel σ-algebra on $\mathbb{R}^2$, denoted as $\mathcal{B}(\mathbb{R}^2)$, is generated by the collection

$$\mathcal{P}_{\mathbb{R}^2} = \{(-\infty, x] \times (-\infty, y] \mid x, y \in \mathbb{R}\}.$$

Thus, we can express this as

$$\mathcal{B}(\mathbb{R}^2) = \sigma(\mathcal{P}_{\mathbb{R}^2}).$$



Theorem 4.8. *Let X and Y be two random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. The mapping $(X(\cdot), Y(\cdot)) : \Omega \to \mathbb{R}^2$ is measurable with respect to $\mathcal{F}$. This means that the pre-images of Borel sets in $\mathbb{R}^2$ under the mapping $(X(\cdot), Y(\cdot))$ correspond to events in the probability space.*

Proof. Let $\mathcal{G}$ be the collection of all subsets of $\mathbb{R}^2$ such that their pre-images under $(X(\cdot), Y(\cdot))$ are events in $\mathcal{F}$. To establish the theorem, it suffices to show that the Borel sets $\mathcal{B}_{\mathbb{R}^2}$ are contained within $\mathcal{G}$.

The set $\mathcal{G}$ is a σ-algebra of subsets of $\mathbb{R}^2$. This can be easily proved just by following through the definitions.

1. **Non-emptiness:**
We first show that $\mathcal{G}$ is non-empty. The empty set $\emptyset$ belongs to $\mathcal{G}$ because its pre-image under any function, including $(X(\cdot), Y(\cdot))$, is the empty set, which is an event in $\mathcal{F}$.

2. **Closed under complements:**
Let $A \in \mathcal{G}$. By definition, the pre-image of A under $(X(\cdot), Y(\cdot))$ is an event in $\mathcal{F}$. We denote this pre-image as $(X, Y)^{-1}(A)$. The complement of A, denoted A^c, has the property that

$$(X, Y)^{-1}(A^c) = \Omega \setminus (X, Y)^{-1}(A),$$

which is also an event in $\mathcal{F}$ since $\mathcal{F}$ is a σ-algebra. Thus, $A^c \in \mathcal{G}$.

3. **Closed under countable unions:**
Let $A_1, A_2, \ldots$ be a countable collection of sets in $\mathcal{G}$. For each n, the pre-image $(X,Y)^{-1}(A_n)$ is an event in $\mathcal{F}$. The pre-image of the union is given by

$$(X,Y)^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} (X,Y)^{-1}(A_n),$$

and since $\mathcal{F}$ is a σ-algebra, this union is also an event in $\mathcal{F}$. Hence, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{G}$.

Since $\mathcal{G}$ satisfies all three properties required for a σ-algebra, we conclude that $\mathcal{G}$ is indeed a σ-algebra.

Next, observe that the sets of the form $\{\omega \mid X(\omega) \leq x\}$ and $\{\omega \mid Y(\omega) \leq y\}$ belong to $\mathcal{F}$ for all $x, y \in \mathbb{R}$ because X and Y are random variables.

Given that $\mathcal{F}$ is a σ-algebra, the intersection of these two sets,

$$\{\omega \mid X(\omega) \leq x\} \cap \{\omega \mid Y(\omega) \leq y\},$$

also belongs to $\mathcal{F}$ for all $x, y \in \mathbb{R}$.

Consequently, we conclude that

$$\{\omega \mid X(\omega) \leq x, Y(\omega) \leq y\} \in \mathcal{F} \text{ for all } x, y \in \mathbb{R}.$$

This implies that the rectangles $(-\infty, x] \times (-\infty, y]$ belong to $\mathcal{G}$ for every $x, y \in \mathbb{R}$ based on the definition of $\mathcal{G}$.

From this, we can see that the collection of sets formed by all finite unions and complements of sets in $\mathcal{G}$ will also belong to $\mathcal{G}$, thus confirming that $\mathcal{G}$ is indeed a σ-algebra.

Finally, we observe that the collection of rectangles $(-\infty, x] \times (-\infty, y]$ generates the Borel σ-algebra $\sigma(\mathcal{P}(\mathbb{R}^2))$, which contains all Borel sets in $\mathbb{R}^2$. Thus, we conclude that $\mathcal{B}_{\mathbb{R}^2} \subseteq \mathcal{G}$. $\square$

Let us denote the space of events as $\mathcal{G}$, where the pre-images of Borel sets on $\mathbb{R}^2$ correspond to events that we can assign probabilities to. This concept leads us to an important definition in probability, known as the *joint probability law.*

Definition 4.12. *The joint probability law for the random variables X and Y is defined as follows:*

$$P_{X,Y}(B) = P(\{\omega \in \Omega \mid (X(\omega), Y(\omega)) \in B\}), \quad B \in \mathcal{B}(\mathbb{R}^2),$$

where $\mathcal{B}(\mathbb{R}^2)$ denotes the Borel σ-algebra on $\mathbb{R}^2$.

This definition captures the idea that we are interested in the probability of the random vector (X, Y) falling within the set B. To illustrate this with a specific example, consider when we take the set B to be $(-\infty, x] \times (-\infty, y]$. In this case, the joint probability law can be expressed as:

$$P_{X,Y}((-\infty, x] \times (-\infty, y]) = P(\{\omega \mid X(\omega) \leq x, Y(\omega) \leq y\}).$$

Here, we are effectively computing the probability that the random variable X takes on a value less than or equal to x and simultaneously, the random variable Y takes on a value less than or equal to y. This joint perspective allows us to understand the relationship between X and Y in a more comprehensive manner.

4.3.1 Joint CDF and Its Properties

Definition 4.13. *Let X and Y be two random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. The joint cumulative distribution function (CDF) of X and Y is defined as:*

$$F_{X,Y}(x,y) = P(\{\omega \mid X(\omega) \leq x, Y(\omega) \leq y\}), \quad \forall x, y \in \mathbb{R}.$$

In simpler terms, we denote this as $F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$.

Properties of Joint CDF

Lemma 4.3. ***Limits at Infinity.*** *We have the following limits:*

$$\lim_{x\to\infty, y\to\infty} F_{X,Y}(x,y) = 1, \quad \textit{and} \quad \lim_{x\to-\infty, y\to-\infty} F_{X,Y}(x,y) = 0.$$

Proof. Consider two unbounded, monotone-increasing sequences $\{x_n\}$ and $\{y_n\}$. Then we can express:

$$\lim_{x\to\infty, y\to\infty} F_{X,Y}(x,y) = \lim_{x\to\infty, y\to\infty} P(X \leq x, Y \leq y) = \lim_{n\to\infty} P(X \leq x_n, Y \leq y_n).$$

By using the properties of probability measures:

$$= P\left(\bigcap_{n=1}^{\infty} \{\omega : X(\omega) \leq x_n, Y(\omega) \leq y_n\}\right) = P(\Omega) = 1.$$

The proof for the second part follows a similar line of reasoning and is left as an exercise to the reader. Notably, the order in which we take the limits does not affect the result. □

Lemma 4.4. ***Monotonicity.*** *For any $x_1 \leq x_2$ and $y_1 \leq y_2$, it holds that:*

$$F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2).$$

Proof. Given $x_1 \leq x_2$ and $y_1 \leq y_2$, the event $\{X \leq x_1, Y \leq y_1\}$ is a subset of the event $\{X \leq x_2, Y \leq y_2\}$. Therefore, we can conclude:

$$P(X \leq x_1, Y \leq y_1) \leq P(X \leq x_2, Y \leq y_2) \Rightarrow F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2).$$

□

Lemma 4.5. ***Continuity from Above.*** *The joint CDF $F_{X,Y}$ is continuous from above:*

$$\lim_{u\to0^+, v\to0^+} F_{X,Y}(x+u, y+v) = F_{X,Y}(x,y), \quad \forall x, y \in \mathbb{R}.$$

Proof. To show that the joint CDF $F_{X,Y}$ is continuous from above, we need to demonstrate that:

$$\lim_{u\to0^+, v\to0^+} F_{X,Y}(x+u, y+v) = F_{X,Y}(x,y) \quad \forall x, y \in \mathbb{R}.$$

consider the expression $F_{X,Y}(x+u, y+v)$:

$$F_{X,Y}(x+u, y+v) = P(X \leq x+u, Y \leq y+v).$$

As $u \to 0^+$ and $v \to 0^+$, the events $\{X \leq x+u\}$ and $\{Y \leq y+v\}$ become increasingly close to the events $\{X \leq x\}$ and $\{Y \leq y\}$.

Specifically, we have:

$$\{X \leq x\} \subseteq \{X \leq x+u\} \quad \text{and} \quad \{Y \leq y\} \subseteq \{Y \leq y+v\}.$$

Therefore, we can express the probability:

$$P(X \leq x+u, Y \leq y+v) \to P(X \leq x, Y \leq y) \quad \text{as } u \to 0^+ \text{ and } v \to 0^+.$$

To formalize this, we can use the continuity of probability measures. For any $\epsilon > 0$, there exist sufficiently small $u > 0$ and $v > 0$ such that:

$$\begin{aligned} P(X \leq x+u, Y \leq y+v) &= P(X \leq x, Y \leq y) + P((X \leq x+u, Y \leq y+v) \setminus (X \leq x, Y \leq y)) \\ &\leq P(X \leq x, Y \leq y) + P(X > x, Y \leq y) + P(X \leq x, Y > y). \end{aligned}$$

As $u \to 0^+$ and $v \to 0^+$, both $P(X > x, Y \leq y)$ and $P(X \leq x, Y > y)$ converge to 0 due to the continuity of the probability measures. Thus, we conclude:

$$\lim_{u \to 0^+, v \to 0^+} F_{X,Y}(x+u, y+v) = F_{X,Y}(x,y).$$

□

Lemma 4.6. ***Marginal CDFs.*** *We also have the property:*

$$\lim_{y \to \infty} F_{X,Y}(x,y) = F_X(x).$$

Proof. Let $\{y_n\}$ be an unbounded, monotone-increasing sequence. Then we can express:

$$\lim_{y \to \infty} F_{X,Y}(x,y) = \lim_{n \to \infty} F_{X,Y}(x, y_n).$$

This leads us to:

$$= \lim_{n \to \infty} P(X \leq x, Y \leq y_n) = P\left(\bigcap_{n=1}^{\infty} \{\omega : X(\omega) \leq x, Y(\omega) \leq y_n\}\right) = P(\{\omega : X(\omega) \leq x\}) = F_X(x),$$

where we again used the continuity of probability measures. □

4.4 Independence of Random Variables

Before we proceed to define the independence of random variables, it is useful to understand the notion of the σ-algebra generated by a random variable.

4.4.1 σ-algebra Generated by Random Variables

We first state an elementary result that holds for any arbitrary function.

Lemma 4.7. *Let Ω and S be two non-empty sets, and let $f : \Omega \to S$ be a function. If H is a σ-algebra of subsets of S, we aim to show that the collection*

$$G = \{A \mid A = f^{-1}(B), B \in H\}$$

is a σ-algebra of subsets of Ω.

In words, the above lemma states that the collection of pre-images of all the sets belonging to some σ-algebra on the range of a function, is a σ-algebra on the domain of that function.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. This random variable X induces a new probability triple $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ on the real line.

Definition 4.14. *We define the σ-algebra generated by the random variable X as follows:*

$$\mathcal{A}(X) = \{E \subseteq \Omega \mid E = X^{-1}(B),\ \forall B \in \mathcal{B}(\mathbb{R})\}.$$

This means that $\mathcal{A}(X)$ consists of all events E in Ω that can be obtained by taking the preimages of Borel sets under the mapping defined by the random variable X.

Lemma 4.8. $\mathcal{A}(X) \subseteq \mathcal{F}$, *which states that the σ-algebra generated by X is indeed a sub-σ-algebra of $\mathcal{F}$.*

To understand this better, note that each Borel set B corresponds to an event E in our probability space. The collection of all such preimages of Borel sets forms the σ-algebra generated by X. Therefore, $\mathcal{A}(X)$ includes exactly those events whose occurrence is fully determined by the value $X(\omega)$ we observe.

Let's illustrate this concept with two examples:

Example 1: Consider a probability space $(\Omega, \mathcal{F}, P)$ and an event $A \in \mathcal{F}$. Define the indicator random variable for the event A as I_A. In this case, we can see that

$$\mathcal{A}(I_A) = \{\emptyset, A, A^c, \Omega\}.$$

Thus, $\mathcal{A}(I_A) \subseteq \mathcal{F}$.

Example 2: Let $([0,1], \mathcal{B}([0,1]), \lambda)$ be our probability space, and consider the random variable $X(\omega) = \omega$ for all $\omega \in \Omega$. In this scenario, we find that

$$\mathcal{A}(X) = \mathcal{F}.$$

From these examples, we observe that the σ-algebra generated by X can either be relatively small (as in Example 1) or as large as the entire σ-algebra $\mathcal{F}$ itself (as in Example 2).

4.4.2 Independence of Random Variables

Definition 4.15. *Two random variables X and Y are independent if the σ-algebras generated by these variables, denoted as $\sigma(X)$ and $\sigma(Y)$, are independent as well.*

To clarify, we can think of independence in terms of events defined by these random variables. Specifically, X and Y are independent if, for any two Borel sets B_1 and B_2 in $\mathbb{R}$, the events corresponding to these sets can be expressed as follows:

$$P\left(\{\omega : X(\omega) \in B_1\} \cap \{\omega : Y(\omega) \in B_2\}\right) = P\left(\{\omega : X(\omega) \in B_1\}\right) \cdot P\left(\{\omega : Y(\omega) \in B_2\}\right)$$

for all $B_1, B_2 \in \mathcal{B}(\mathbb{R})$.

This means that the probability of both X falling within B_1 and Y falling within B_2 simultaneously is simply the product of their individual probabilities of falling within these sets.

Additionally, there exists a helpful theorem that characterizes the independence of random variables in terms of their joint cumulative distribution function (CDF).

Theorem 4.9. *X and Y are independent if and only if the joint CDF of X and Y can be expressed as the product of their individual marginal CDFs.*

In mathematical terms, if $F_{X,Y}(x,y)$ represents the joint CDF of X and Y, and $F_X(x)$ and $F_Y(y)$ denote the marginal CDFs, then X and Y are independent if:

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$$

for all x, y. This equivalence provides a practical method for checking the independence of random variables based on their distribution functions.

Proof. To prove this theorem, we need to establish both the necessary and sufficient conditions for the independence of random variables X and Y.

Proof of the necessary condition:

Assume that X and Y are independent random variables. By the definition of independence, for any two Borel sets $B_1 \in \mathcal{B}(\mathbb{R})$ and $B_2 \in \mathcal{B}(\mathbb{R})$, the events $\{\omega \mid X(\omega) \in B_1\}$ and $\{\omega \mid Y(\omega) \in B_2\}$ are independent. This means that:

$$P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2).$$

From the definition of the joint cumulative distribution function (CDF), we have:

$$P(X \in B_1, Y \in B_2) = P_{X,Y}(B_1 \times B_2).$$

Thus, we can write:

$$P_{X,Y}(B_1 \times B_2) = P(X \in B_1)P(Y \in B_2).$$

This holds for all Borel sets in $\mathbb{R}$.

Now, we specifically choose the sets $B_1 = (-\infty, x]$ and $B_2 = (-\infty, y]$. Consequently, we obtain:

$$F_{X,Y}(x,y) = P(X \le x, Y \le y) = P(X \le x)P(Y \le y) = F_X(x)F_Y(y),$$

for all $x, y \in \mathbb{R}$. This completes the proof of the necessary condition.

Proof of the sufficient condition:

Now, we will prove the sufficiency part. Assume that the joint cumulative distribution function (CDF) of X and Y can be expressed as:

$$F_{X,Y}(x,y) = F_X(x)F_Y(y), \quad \forall x, y \in \mathbb{R}.$$

To show that X and Y are independent, we need to demonstrate that for any Borel sets B_1 and B_2, the events defined by these sets are independent.

Using the property of joint CDFs, we can express the probability of the joint event as follows:

$$P(X \in B_1, Y \in B_2) = P(X \le x_1, Y \le y_1) \quad \text{for } (x_1, y_1) \in B_1 \times B_2.$$

Using the given expression for the joint CDF, we have:

$$P(X \in B_1, Y \in B_2) = F_{X,Y}(x_1, y_1) = F_X(x_1)F_Y(y_1).$$

Now, we also know:

$$P(X \in B_1) = F_X(x_1) \quad \text{and} \quad P(Y \in B_2) = F_Y(y_1).$$

Thus, we can write:

$$P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2).$$

Since this holds for any arbitrary Borel sets B_1 and B_2, we conclude that the events $\{X \in B_1\}$ and $\{Y \in B_2\}$ are independent.

Therefore, we have established that if the joint CDF can be expressed as the product of the marginal CDFs, then X and Y are indeed independent.

$\square$

Let's extend our results to a general n random variables.

Definition 4.16. *Random variables $X_1, X_2, \ldots, X_n$ are independent if the collections of events associated with these variables, represented by their σ-algebras $\sigma(X_1), \sigma(X_2), \ldots, \sigma(X_n)$, are independent.*

This means that for any measurable sets $B_i \in \mathcal{B}(\mathbb{R})$ for $1 \leq i \leq n$, the joint probability can be expressed as the product of their individual probabilities. Mathematically, this is written as:

$$P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i).$$

In simpler terms, knowing the outcome of one random variable does not give us any information about the others when they are independent.

Furthermore, we can characterize the independence of these random variables in terms of their cumulative distribution functions (CDFs).

Theorem 4.10. *The random variables $X_1, X_2, \ldots, X_n$ are independent if and only if their joint CDF can be expressed as the product of their individual CDFs.*

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i).$$

This theorem emphasizes that the joint distribution function of the independent random variables is completely determined by their individual distribution functions. In essence, this means that the behavior of the entire system of random variables can be understood by examining each variable independently, highlighting the fundamental nature of independence in probability.

Definition 4.17. *A family of random variables $\{X_i\}_{i \in I}$ as independent if the collections of events associated with these random variables, represented by their σ-algebras $\{\sigma(X_i) : i \in I\}$, are independent.*

This means that for any finite selection of indices $i_1, i_2, \ldots, i_n$ from the index set I, the joint probability of the corresponding random variables can be expressed as the product of their individual probabilities. Mathematically, for any sets $B_{i_j} \in \mathcal{B}(\mathbb{R})$ for $j = 1, 2, \ldots, n$, we have:

$$P(X_{i_1} \in B_{i_1}, X_{i_2} \in B_{i_2}, \ldots, X_{i_n} \in B_{i_n}) = \prod_{j=1}^{n} P(X_{i_j} \in B_{i_j}).$$

In simpler words, the independence of the family of random variables implies that knowing the value of one variable provides no information about the values of the others. Thus, the concept of independence can be generalized beyond a finite number of random variables to an arbitrary family indexed by a set I, reinforcing the idea that the individual behavior of each variable remains unaffected by the others in the family.

4.5 Conditional Distributions and Joint Continuity

4.5.1 Joint PMF of Discrete Random Variables

If X and Y are discrete random variables, the range of the mapping $(X(\cdot), Y(\cdot))$ forms a countable subset of $\mathbb{R}^2$. This arises from the fact that the Cartesian product of two countable sets is also countable. To clarify, a set is considered countable if it can be put into a one-to-one correspondence with the natural numbers. Therefore, since both X and Y take on countably many values, their combined behavior represented by (X, Y) will also be countable, thus making $(X(\cdot), Y(\cdot))$ a discrete random variable on $\mathbb{R}^2$.

It's important to note that we cannot make the same assertion when X and Y are continuous random variables. In such cases, even if X and Y are each continuous on their own, they may not be jointly continuous. Refer subsection 4.5.3 for this.

The joint probability mass function (pmf) of the discrete random variables X and Y is given by:

$$p_{X,Y}(x, y) = P(X = x, Y = y), \quad x, y \in \mathbb{R}.$$

This joint pmf plays a crucial role in defining the joint distribution of X and Y. Specifically, for any Borel set $B \in \mathcal{B}_{\mathbb{R}^2}$, the probability that the random vector (X, Y) falls within the set B is given by:

$$P_{X,Y}(B) = \sum_{(x,y)\in B} p_{X,Y}(x, y).$$

This equation shows that the joint pmf allows us to calculate probabilities associated with any region in the plane defined by the variables X and Y.

4.5.2 Conditional PMF of Discrete Random Variables

Now, we define the conditional probability mass function (pmf) for discrete random variables.

Definition 4.18. *Let X and Y be discrete random variables defined on the probability space $(\Omega, \mathcal{F}, P)$. The conditional probability of X given Y is defined as:*

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

where $p_Y(y) > 0$.

The following theorem describes the concept of independence for discrete random variables in terms of the conditional pmf.

Theorem 4.11. *The following statements are equivalent for discrete random variables X and Y:*

1. *X and Y are independent.*
2. *For all $x, y \in \mathbb{R}$, the events $\{X = x\}$ and $\{Y = y\}$ are independent.*
3. *For all $x, y \in \mathbb{R}$, $P_{X,Y}(x, y) = P_X(x)P_Y(y)$.*
4. *For all $x, y \in \mathbb{R}$ such that $p_Y(y) > 0$, $p_{X|Y}(x|y) = p_X(x)$. ''*

Proof. The equivalences (2) ⇔ (3) and (3) ⇔ (4) follow directly from the definitions of independence and the conditional pmf.

Now, let us prove the equivalence between (1) and (3):

(1) implies (3): If X and Y are independent, then the joint probability of X and Y can be expressed as:

$$P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2).$$

Now, let $B_1 = \{x\}$ and $B_2 = \{y\}$. Therefore, we have:

$$P(X = x, Y = y) = P(X = x)P(Y = y).$$

This establishes statement (3).

(3) implies (1): Given that $P(X \in B_1, Y \in B_2) = P_{X,Y}(x, y)$, we can write:

$$P(X \in B_1, Y \in B_2) = \sum_{x \in B_1} \sum_{y \in B_2} p_{X,Y}(x, y).$$

Using statement (3), we substitute to obtain:

$$\sum_{x \in B_1} \sum_{y \in B_2} P_X(x)P_Y(y) = \sum_{x \in B_1} P_X(x) \sum_{y \in B_2} P_Y(y) = P(X \in B_1)P(Y \in B_2).$$

This confirms statement (1). Thus, we have shown that the statements are equivalent. □

4.5.3 Joint PMF of Continuous Random Variables

Definition 4.19. *Two random variables X and Y are termed jointly continuous if their joint probability distribution, denoted $P_{X,Y}$, is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$. This means that for every Borel set $N \subset \mathbb{R}^2$ with Lebesgue measure zero, we have:*

$$P((X, Y) \in N) = 0.$$

In simple terms, this ensures that there is no *mass* of probability concentrated on any set of points with zero area in $\mathbb{R}^2$. For X and Y to be jointly continuous, their probability distribution must spread smoothly over regions in $\mathbb{R}^2$, without *spikes* or *isolated points.*

The *Radon-Nikodym Theorem* gives us a more practical tool here.

Theorem 4.12. *X and Y are jointly continuous random variables if and only if there exists a measurable function $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$, called the joint probability density function (pdf), such that for any region $B \subset \mathbb{R}^2$,*

$$P((X, Y) \in B) = \int_B f_{X,Y}(u, v)\, d\lambda(u, v),$$

where λ represents the Lebesgue measure on $\mathbb{R}^2$.

Implication of Joint Continuity: To express this in a familiar form, let's examine the probability that X and Y fall within certain ranges. For $B = (-\infty, x] \times (-\infty, y]$, this probability becomes:

$$F_{X,Y}(x, y) = P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, dv\, du,$$

where $F_{X,Y}(x, y)$ is the *joint cumulative distribution function* (cdf), and $f_{X,Y}(x, y)$ is the *joint pdf.*

Thus, the joint pdf $f_{X,Y}(x, y)$ provides a complete description of the joint behavior of X and Y. This means that knowing $f_{X,Y}(x, y)$ allows us to compute the probabilities for any event involving X and Y over regions in $\mathbb{R}^2$.

Lemma 4.9. *If X is continuous and Y is continuous, then (X, Y) need not be jointly continuous.*

We won't formally prove this lemma, but we will provide an example that confirms that this is true.

Example 4.1. *Suppose $X \sim N(0,1)$, meaning that X is a standard normal random variable with mean 0 and variance 1, and $Y = 2X$. So, Y will have a mean of 0 and a variance of 4, which we can verify by noting that*

$$Y \sim N(0,4).$$

Even though both X and Y are continuous random variables, we need to examine if they form a jointly continuous pair when taken together as the random vector (X, Y).

A pair of random variables (X, Y) is considered jointly continuous if there exists a joint probability density function $f_{X,Y}(x, y)$ over the entire $\mathbb{R}^2$ space, such that for any subset $A \subset \mathbb{R}^2$,

$$P((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy.$$

Notice that $Y = 2X$ directly ties Y to X. This means Y is perfectly linearly dependent on X, implying that (X, Y) does not vary freely across the $\mathbb{R}^2$ plane. Instead, the values of (X, Y) are restricted to a specific line, namely $y = 2x$. Since $Y = 2X$, the pair (X, Y) essentially lies along the line $y = 2x$ in the $\mathbb{R}^2$ plane. Consequently, the joint probability measure of (X, Y) is concentrated along this line rather than spread across two dimensions.

In terms of probability density, this restriction means that there does not exist a two-dimensional joint density $f_{X,Y}(x, y)$ over the entire $\mathbb{R}^2$ space, because we cannot assign probabilities over regions in $\mathbb{R}^2$ outside this line. Any density that would describe (X, Y) is confined to one dimension along $y = 2x$.

Thus, although X and Y are individually continuous, (X, Y) as a pair does not possess joint continuity in two dimensions.

Lemma 4.10. *When X and Y are jointly continuous random variables, their marginal distributions are also continuous. To understand why, let's examine the probability that $X \leq x$ and $Y \leq y$:*

$$P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du$$

Now, we want to find the marginal distribution of X. The marginal probability $P(X \leq x)$ is obtained by integrating over all possible values of Y:

$$P(X \leq x) = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f_{X,Y}(u,v)\,dv \right) du = \int_{-\infty}^{x} f_X(u)\,du$$

Here, the expression inside the parentheses,

$$\int_{-\infty}^{\infty} f_{X,Y}(u,v)\,dv,$$

produces a function of u which is non-negative and measurable. This function represents the density of X, which we denote as $f_X(u)$.

Thus, we can write:

$$f_X(u) = \int_{-\infty}^{\infty} f_{X,Y}(u,v)\,dv$$

This shows that X has a continuous marginal distribution with f_X as its probability density function (pdf). Since $f_X(u)$ is derived as an integral over a non-negative function $f_{X,Y}(u,v)$, it ensures that f_X itself is also non-negative and measurable.

A similar argument applies to the marginal pdf of Y, confirming that both X and Y are indeed continuous. This continuity of the marginals follows directly from the joint continuity of X and Y, as we have shown through their integrals.

Definition 4.20. ***Independence of Joint Continuous Random Variables.*** *For two random variables X and Y, we say they are* ***independent*** *if and only if their joint distribution function $F_{X,Y}(x,y)$ can be written as the product of their marginal distribution functions. In other words,*

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \forall x, y \in \mathbb{R}.$$

Now, let's apply this definition to a specific case: when X and Y are **jointly continuous** random variables. For continuous variables, we express independence in terms of their probability density functions. That is, we integrate the joint density function $f_{X,Y}(x,y)$ over the range $(-\infty, x]$ and $(-\infty, y]$:

$$\int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du = \left(\int_{-\infty}^{x} f_X(u)\,du\right)\left(\int_{-\infty}^{y} f_Y(v)\,dv\right).$$

Expanding this, we see that independence requires

$$\int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\,dv\,du = \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u)f_Y(v)\,dv\,du.$$

Since this equality holds for *all* values of x and y, we can conclude that the two integrands themselves must be equal almost everywhere. Therefore,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \forall x, y \in \mathbb{R},$$

except possibly on a subset of $\mathbb{R}^2$ with Lebesgue measure zero. This condition—that the joint density factorizes as the product of the marginal densities—is both a **necessary and sufficient condition** for the independence of two jointly continuous random variables.

4.5.4 Conditional PDF of Continuous Random Variables

To define the conditional cumulative distribution function (CDF) $F_{X|Y}(x|y) \approx P(X \leq x | Y = y)$, we face a technical challenge: the event $\{Y = y\}$ has probability zero for any specific y when Y is a continuous random variable. This zero-probability event makes a direct conditional probability definition problematic.

To resolve this, we approximate by conditioning on Y taking values within a small interval $(y, y+\epsilon)$, where $\epsilon > 0$. We then examine what happens as ϵ approaches zero. This approach leads to a motivating derivation for defining the conditional probability density function (pdf) in the jointly continuous case.

The conditional CDF of X, given that Y is *close to y*, can be approximated by:

$$F_{X|Y}(x|y) = P(X \leq x \mid y \leq Y \leq y + \epsilon) \quad \text{(for small } \epsilon\text{)}.$$

This probability can be rewritten in terms of joint and marginal probabilities as:

$$F_{X|Y}(x|y) = \frac{P(\{X \leq x\} \cap \{y \leq Y \leq y+\epsilon\})}{P(y \leq Y \leq y+\epsilon)}.$$

To express this in terms of CDFs, we observe that:

$$F_{X|Y}(x|y) = \frac{F_{X,Y}(x, y+\epsilon) - F_{X,Y}(x, y)}{F_Y(y+\epsilon) - F_Y(y)}.$$

Now, dividing the numerator and denominator by ϵ, we obtain:

$$F_{X|Y}(x|y) = \frac{\frac{F_{X,Y}(x,y+\epsilon) - F_{X,Y}(x,y)}{\epsilon}}{\frac{F_Y(y+\epsilon) - F_Y(y)}{\epsilon}}.$$

As $\epsilon \to 0$, the right-hand side of this expression resembles the form of a derivative. Specifically, it suggests:

$$F_{X|Y}(x|y) = \frac{\frac{\partial}{\partial y} F_{X,Y}(x, y)}{\frac{d}{dy} F_Y(y)}.$$

This expression provides a foundation for defining the conditional CDF and, by extension, the conditional PDF.

Definition 4.21. *The* ***conditional cumulative distribution function*** *(CDF) of a random variable X given $Y = y$ is a way to accumulate probability up to a certain point x, while taking into account the information provided by $Y = y$. It is defined as:*

$$F_{X|Y}(x|y) = \int_{-\infty}^{x} \frac{f_{X,Y}(u, y)}{f_Y(y)}\, du,$$

where $f_{X,Y}(u, y)$ is the joint probability density function (pdf) of X and Y, and $f_Y(y)$ is the marginal pdf of Y.

Note that this expression is valid as long as $f_Y(y) > 0$, since we cannot condition on an impossible or zero-probability event.

Definition 4.22. *The* ***conditional probability density function*** *(pdf) of X given $Y = y$ tells us the probability density of X at a particular point x, assuming $Y = y$. It is defined as:*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)},$$

again, assuming $f_Y(y) > 0$.

This expression gives us a way to interpret X as a function of the specific value y of Y, concentrating our attention on the values X takes when $Y = y$.

Definition 4.23. *Suppose we have an event A that belongs to the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ (denoted $A \in \mathcal{B}(\mathbb{R})$, the collection of all Borel subsets of $\mathbb{R}$). The* ***conditional probability*** *of the event A given $Y = y$ is defined as:*

$$P(X \in A | Y = y) = \int_A f_{X|Y}(v|y)\, dv = \int_{-\infty}^{\infty} \mathbb{I}_A(v) f_{X|Y}(v|y)\, dv,$$

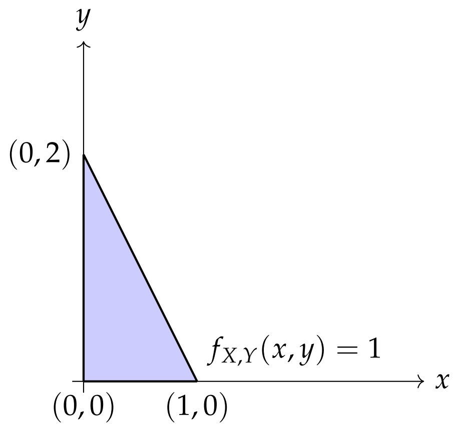
where $\mathbb{I}_A(v)$ is the indicator function for the set A, which is 1 if $v \in A$ and 0 otherwise.

This integral sums up the conditional pdf $f_{X|Y}(v|y)$ over the region defined by A, thereby giving the probability that X lies in A, under the condition that $Y = y$.

Exercise 4.8. *Consider random variables X and Y with a joint continuous distribution defined over a triangular region with vertices at points $(0,0)$, $(0,2)$, and $(1,0)$. The joint probability density function is given by:*

$$f_{X,Y}(x,y) = 1$$

for (x,y) within the triangular region. Find the marginal distributions $f_X(x)$ and $f_Y(y)$, and the conditional distributions $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$.



Solution 4.8. *The region where $f_{X,Y}(x,y) = 1$ is the triangle with vertices $(0,0)$, $(0,2)$, and $(1,0)$. The line connecting $(0,2)$ and $(1,0)$ has the equation:*

$$y = 2 - 2x$$

Therefore, the joint distribution $f_{X,Y}(x,y) = 1$ is defined for $0 \le x \le 1$ and $0 \le y \le 2 - 2x$.

To ensure $f_{X,Y}(x,y) = 1$ is a valid probability density function, we need:

$$\int_0^1 \int_0^{2-2x} 1\,dy\,dx = 1$$

Compute this integral:

$$\int_0^1 \int_0^{2-2x} 1\,dy\,dx = \int_0^1 (2-2x)\,dx = \int_0^1 2 - 2x\,dx = \left[2x - x^2\right]_0^1 = 2 - 1 = 1$$

Thus, $f_{X,Y}(x,y) = 1$ is indeed normalized.

To find $f_X(x)$, integrate $f_{X,Y}(x,y)$ over y:

$$f_X(x) = \int_0^{2-2x} 1\,dy = 2 - 2x$$

for $0 \le x \le 1$.

To find $f_Y(y)$, integrate $f_{X,Y}(x,y)$ over x. Note that for a given y, x ranges from 0 to $1 - \frac{y}{2}$ (from the line equation rearranged):

$$f_Y(y) = \int_0^{1-\frac{y}{2}} 1\,dx = 1 - \frac{y}{2}$$

for $0 \le y \le 2$.

The conditional distribution $f_{Y|X}(y|x)$ *is given by:*

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{1}{2-2x}$$

for $0 \leq y \leq 2-2x$.

The conditional distribution $f_{X|Y}(x|y)$ *is given by:*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{1-\frac{y}{2}}$$

for $0 \leq x \leq 1-\frac{y}{2}$.

Exercise 4.9. *Two persons X and Y live in cities A and B but work in cities B and A respectively. Every morning they start for work at a uniformly random time between 9 am and 10 am independent of each other. Both of them travel at the same constant speed and it takes 20 minutes to reach the other city. What is the probability that X and Y meet each other on their way to work?*

Solution 4.9. *Let* T_X *and* T_Y *be the times at which persons X and Y leave their respective cities. Both* T_X *and* T_Y *are uniformly distributed over the interval* $[0, 60]$ *minutes, where 0 corresponds to 9:00 am and 60 corresponds to 10:00 am. The travel time for both X and Y is 20 minutes. Therefore X will reach city B at* $T_X + 20$ *minutes and Y will reach city A at* $T_Y + 20$ *minutes.*

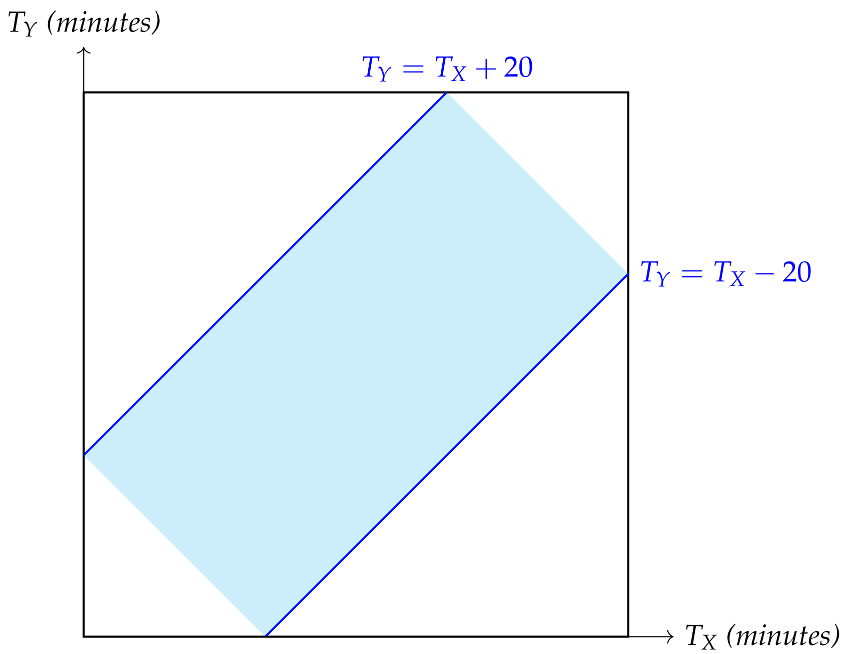
For X and Y to meet, they must be on the road at the same time. This occurs if X leaves before Y arrives in city A, and Y leaves before X arrives in city B, which can be expressed as:

$$T_X + 20 > T_Y \quad \textit{and} \quad T_Y + 20 > T_X.$$

These inequalities can be rearranged to:

$$T_Y < T_X + 20 \quad \textit{and} \quad T_X < T_Y + 20.$$

This forms a square on the coordinate plane where the side length is 60 minutes (the range of the departure times).



1. The area of the square representing all possible combinations of departure times is:

$$A_{total} = 60 \times 60 = 3600.$$

2. The region where they meet can be represented by the inequalities. Graphically, this creates a band around the line $T_Y = T_X$ within the square. The bounds for T_Y are:

$$T_Y < T_X + 20 \quad \textit{and} \quad T_Y > T_X - 20.$$

3. This forms a parallelogram with vertices at (0, 20), (40, 60), (60, 40), and (20, 0). The area of this meeting region can be calculated by:

$$A_{meet} = 60^2 - 4 \times \frac{1}{2} \times 20 \times 20 = 3600 - 800 = 2800.$$

4. Finally, the probability P that X and Y meet each other on their way to work is given by the ratio of the area where they meet to the total area:

$$P = \frac{A_{meet}}{A_{total}} = \frac{2800}{3600} = \frac{7}{9}.$$

Exercise 4.10. *Data is taken on the height and shoe size of a sample of MIT students. Height (X) is coded by 3 values: 1 (short), 2 (average), 3 (tall) and Shoe size (Y) is coded by 3 values: 1 (small), 2 (average), 3 (large). The joint counts are given in the following table:*

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 1$	234	225	84
$X = 2$	180	453	161
$X = 3$	39	192	157

(a) Find the joint and marginal pmf of X and Y.
(b) Are X and Y independent? Discuss in detail.

Solution 4.10. *To solve the problem, we start by calculating the joint probability mass function (pmf) of X and Y from the given joint counts.*

Let N be the total number of observations, which is the sum of all joint counts:

$$N = 234 + 225 + 84 + 180 + 453 + 161 + 39 + 192 + 157 = 1260.$$

The joint probabilities $P(X = x, Y = y)$ can be calculated as follows:

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 1$	$\frac{234}{1260}$	$\frac{225}{1260}$	$\frac{84}{1260}$
$X = 2$	$\frac{180}{1260}$	$\frac{453}{1260}$	$\frac{161}{1260}$
$X = 3$	$\frac{39}{1260}$	$\frac{192}{1260}$	$\frac{157}{1260}$

Next, we find the marginal pmfs $P(X = x)$ and $P(Y = y)$.

For $P(X = x)$:

$$\begin{aligned}
P(X=1) &= P(X=1,Y=1)+P(X=1,Y=2)+P(X=1,Y=3) \\
&= \frac{234+225+84}{1260} = \frac{543}{1260} = 0.4300, \\
P(X=2) &= P(X=2,Y=1)+P(X=2,Y=2)+P(X=2,Y=3) \\
&= \frac{180+453+161}{1260} = \frac{794}{1260} = 0.6286, \\
P(X=3) &= P(X=3,Y=1)+P(X=3,Y=2)+P(X=3,Y=3) \\
&= \frac{39+192+157}{1260} = \frac{388}{1260} = 0.3080.
\end{aligned}$$

For $P(Y = y)$:

$$\begin{aligned}
P(Y=1) &= P(X=1,Y=1)+P(X=2,Y=1)+P(X=3,Y=1) \\
&= \frac{234+180+39}{1260} = \frac{453}{1260} = 0.3596, \\
P(Y=2) &= P(X=1,Y=2)+P(X=2,Y=2)+P(X=3,Y=2) \\
&= \frac{225+453+192}{1260} = \frac{870}{1260} = 0.6905, \\
P(Y=3) &= P(X=1,Y=3)+P(X=2,Y=3)+P(X=3,Y=3) \\
&= \frac{84+161+157}{1260} = \frac{402}{1260} = 0.3183.
\end{aligned}$$

Now, to check for independence, X and Y are independent if:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \quad \textit{for all } x, y.$$

Calculating for $P(X = 1)$ and $P(Y = 1)$:

$$P(X = 1) \cdot P(Y = 1) = 0.4300 \cdot 0.3596 \approx 0.1540.$$

However, we find:

$$P(X = 1, Y = 1) = \frac{234}{1260} \approx 0.1857.$$

Since $0.1857 \neq 0.1540$, X and Y are not independent.

Exercise 4.11. *John is vacationing in Monte Carlo. Each evening, the amount of money he takes to the casino is a random variable X with the pdf*

$$f_X(x) = \begin{cases} Cx & 0 < x \leq 100 \\ 0 & \textit{elsewhere} \end{cases}$$

At the end of each night, the amount Y he returns with is uniformly distributed between zero and twice the amount he came to the casino with.

(a) Find the value of C.

(b) For a fixed α, $0 \leq \alpha \leq 100$, what is the conditional pdf of Y given $X = \alpha$?

(c) If John goes to the casino with α dollars, what is the probability he returns with more than α dollars?

(d) *Determine the joint pdf, $f_{X,Y}(x,y)$, of X and Y as well as the marginal pdf, $f_Y(y)$, of Y.*

Solution 4.11. *To solve the problem, we will address each part sequentially.*

(a) *To find the value of C, we use the property that the total area under the pdf must equal 1:*

$$\int_0^{100} Cx\,dx = 1.$$

Evaluating the integral:

$$\int_0^{100} Cx\,dx = C\left[\frac{x^2}{2}\right]_0^{100} = C \cdot \frac{100^2}{2} = 5000C.$$

Setting this equal to 1 gives:

$$5000C = 1 \implies C = \frac{1}{5000}.$$

(b) *The conditional pdf of Y given $X = \alpha$ is uniform over the interval $[0, 2\alpha]$:*

$$f_{Y|X}(y|\alpha) = \begin{cases} \frac{1}{2\alpha} & 0 \le y \le 2\alpha \\ 0 & \textit{elsewhere} \end{cases}$$

(c) *To find the probability that John returns with more than α dollars given he started with α, we need:*

$$P(Y > \alpha | X = \alpha) = \int_\alpha^{2\alpha} f_{Y|X}(y|\alpha)\,dy = \int_\alpha^{2\alpha} \frac{1}{2\alpha}\,dy = \frac{1}{2\alpha}\,[y]_\alpha^{2\alpha} = \frac{1}{2\alpha}(2\alpha - \alpha) = \frac{1}{2}.$$

(d) *The joint pdf $f_{X,Y}(x,y)$ can be expressed as:*

$$f_{X,Y}(x,y) = f_X(x) \cdot f_{Y|X}(y|x) = \begin{cases} Cx \cdot \frac{1}{2x} & 0 < x \le 100, 0 < y \le 2x \\ 0 & \textit{elsewhere} \end{cases}$$

This simplifies to:

$$f_{X,Y}(x,y) = \begin{cases} \frac{C}{2}x & 0 < x \le 100, 0 < y \le 2x \\ 0 & \textit{elsewhere} \end{cases}$$

The marginal pdf of Y is found by integrating the joint pdf:

$$f_Y(y) = \int_0^{100} f_{X,Y}(x,y)\,dx.$$

The limits for x depend on y; specifically, y must satisfy $0 < y \le 2x \implies x \ge \frac{y}{2}$. Thus, we integrate:

$$f_Y(y) = \int_{\frac{y}{2}}^{100} \frac{C}{2}x\,dx.$$

Calculating this integral:

$$f_Y(y) = \frac{C}{2}\left[\frac{x^2}{2}\right]_{\frac{y}{2}}^{100} = \frac{C}{2}\left(\frac{100^2}{2} - \frac{\left(\frac{y}{2}\right)^2}{2}\right) = \frac{C}{2}\left(5000 - \frac{y^2}{8}\right).$$

Thus,

$$f_Y(y) = \begin{cases} \frac{C}{2}\left(5000 - \frac{y^2}{8}\right) & 0 < y \le 200 \\ 0 & \textit{elsewhere} \end{cases}$$

Exercise 4.12. *A rod is broken at two points that are chosen uniformly and independently at random. What is the probability that the three resulting pieces form a triangle?*

Solution 4.12. *To determine the probability that the three resulting pieces from breaking the rod can form a triangle, we can apply the triangle inequality. For three lengths a, b, and c to form a triangle, the following conditions must hold:*

$$a+b>c, \quad a+c>b, \quad b+c>a$$

Let the length of the rod be L. When the rod is broken at two points, say X and Y, chosen uniformly at random along its length, we can denote the positions of the breaks as X and Y, where $0<X<Y<L$*. This creates three segments of lengths:*

$$a=X, \quad b=Y-X, \quad c=L-Y$$

The inequalities for these segments can be expressed as:

1. $X+(Y-X)>(L-Y)$ *which simplifies to* $Y>\frac{L}{2}$
2. $X+(L-Y)>(Y-X)$ *which simplifies to* $X+L-Y>Y-X$ *or* $2X+L>2Y$ *which can be rewritten as* $2X>2Y-L$
3. $(Y-X)+(L-Y)>X$ *which simplifies to* $L-X>X$ *or* $L>2X$ *giving* $X<\frac{L}{2}$

To visualize this, consider the unit square $[0,L]\times[0,L]$ *where the points* (X,Y) *fall within the bounds* $0<X<Y<L$*. We are interested in the region where all three inequalities are satisfied.*

The area of the valid region can be determined as follows:
1. *The condition* $Y>\frac{L}{2}$ *restricts Y to the upper half of the square for* $X<Y$*.*
2. *The condition* $X<\frac{L}{2}$ *restricts X to the left half of the square.*
3. *The condition* $2X>2Y-L$ *introduces a linear constraint.*

By determining the intersections of these regions within the unit square, we can find the area where all conditions are satisfied. (Try it out yourself!)

The calculations yield that the area where the pieces can form a triangle is $\frac{1}{4}$ *of the total area of the triangle formed by the breaks.*

Exercise 4.13. *Melvin Fooch, a student of probability theory, has found that the hours he spends working (W) and sleeping (S) in preparation for a final exam are random variables described by:*

$$f_{W,S}(w,s)=\begin{cases} K, & 10\le w+s\le 20 \text{ and } w\ge 0, s\ge 0 \\ 0, & \text{elsewhere.} \end{cases}$$

What poor Melvin does not know, and even his best friends will not tell him, is that working only furthers his confusion and that his grade, G, can be described by

$$G=2.5(S-W)+50.$$

(a) *The instructor has decided to pass Melvin if, on the exam, he achieves* $G\ge 75$*. What is the probability that this will occur?*

(b) *Suppose Melvin got a grade greater than or equal to 75 on the exam. Determine the conditional probability that he spent less than one hour working in preparation for this exam.*

(c) *Are the random variables W and S independent? Justify.*

Solution 4.13. *To solve the exercise, we first need to find the value of K such that the joint probability density function $f_{W,S}(w,s)$ integrates to 1 over the valid range.*

The region defined by $10 \leq w+s \leq 20$ can be represented in the $w-s$ plane. The limits are $w+s=10$ and $w+s=20$.

The area of interest is a trapezoid with vertices at $(0,10),(10,0),(20,0),(0,20)$.

The area can be calculated as:

$$A = \textit{Area of trapezoid} = \frac{1}{2} \times (b_1+b_2) \times h = \frac{1}{2} \times (10+20) \times 10 = 150$$

where $b_1 = 10$, $b_2 = 20$, and $h = 10$.

Since the total probability must equal 1:

$$K \times 150 = 1 \implies K = \frac{1}{150}$$

The grade G can be expressed as:

$$G \geq 75 \implies 2.5(S-W)+50 \geq 75 \implies S-W \geq 10 \implies S \geq W+10$$

We now need to find the area of the region where $S \geq W+10$ within the trapezoid $10 \leq w+s \leq 20$.

The line $S = W+10$ intersects $w+s=20$ at $(10,10)$ and does not exceed $w+s=10$.

The area under consideration is now a triangle with vertices $(0,10),(10,10),(10,0)$.

Area of this triangle is:

$$A_{pass} = \frac{1}{2} \times 10 \times 10 = 50$$

Thus, the probability that Melvin passes is:

$$P(G \geq 75) = K \cdot A_{pass} = \frac{1}{150} \cdot 50 = \frac{1}{3}.$$

To find the conditional probability $P(W<1|G \geq 75)$, we need the area where $W<1$ and $S \geq W+10$ within the region where $G \geq 75$.

The line $W=1$ intersects $S=1+10=11$. The area for $W<1$ is a triangle with vertices $(0,10),(1,11),(1,9)$.

Area of this triangle:

$$A_{conditional} = \frac{1}{2} \times 1 \times 1 = \frac{1}{2}.$$

Now, the conditional probability is:

$$P(W<1|G \geq 75) = \frac{A_{conditional}}{A_{pass}} = \frac{\frac{1}{2}}{50} = \frac{1}{100}.$$

Random variables W and S are independent if and only if $f_{W,S}(w,s) = f_W(w) \cdot f_S(s)$.

We can find $f_W(w)$ and $f_S(s)$ by integrating $f_{W,S}(w,s)$:

$$f_W(w) = \int_0^{20-w} f_{W,S}(w,s)\,ds = \int_0^{20-w} K\,ds = K(20-w).$$

$$f_S(s) = \int_0^{20-s} f_{W,S}(w,s)\,dw = \int_0^{20-s} K\,dw = K(20-s).$$

Thus, the marginal distribution for S is given by:

$$f_S(s) = \begin{cases} K(20-s), & 0 \le s \le 20 \\ 0, & \textit{otherwise.} \end{cases}$$

Now, we can compute the product $f_W(w) \cdot f_S(s)$:

$$f_W(w) \cdot f_S(s) = (K(20-w)) \cdot (K(20-s)) = K^2(20-w)(20-s).$$

To check for independence, we compare this product to the joint distribution $f_{W,S}(w,s)$:

$$f_{W,S}(w,s) = K \quad \textit{for } 10 \le w+s \le 20 \textit{ and } w \ge 0, s \ge 0.$$

Since $K^2(20-w)(20-s)$ is not equal to K for all w and s in the defined region (specifically, when both w and s take on values such that $w+s$ is between 10 and 20), the product $f_W(w) \cdot f_S(s)$ does not equal $f_{W,S}(w,s)$.

Therefore, we conclude that the random variables W and S are not independent.

Exercise 4.14. *Stations A and B are connected by two parallel message channels. A message from A to B is sent over both channels at the same time. Continuous random variables X and Y represent the message delays (in hours) over parallel channels I and II, respectively. These two random variables are independent, and both are uniformly distributed from 0 to 1 hours. A message is considered received as soon as it arrives on any one channel, and it is considered verified as soon as it has arrived over both channels.*

(a) Determine the probability that a message is received within 15 minutes after it is sent.

(b) Determine the probability that the message is received but not verified within 15 minutes after it is sent.

(c) If the attendant at B goes home 15 minutes after the message is received, what is the probability that he is present when the message should be verified?

Solution 4.14. *Let us first convert the time of 15 minutes into hours:*

$$15 \textit{ minutes} = \frac{15}{60} = 0.25 \textit{ hours.}$$

A message is received if at least one of the channels has a delay less than or equal to 0.25 hours. The probability that $X > 0.25$ and $Y > 0.25$ gives us the probability that the message is not received:

$$P(X > 0.25) = 1 - P(X \le 0.25) = 1 - 0.25 = 0.75,$$

$$P(Y > 0.25) = 1 - P(Y \le 0.25) = 1 - 0.25 = 0.75.$$

Since X and Y are independent,

$$P(X > 0.25 \textit{ and } Y > 0.25) = P(X > 0.25) \cdot P(Y > 0.25) = 0.75 \cdot 0.75 = 0.5625.$$

Thus, the probability that a message is received within 15 minutes is:

$$P(received) = 1 - P(X > 0.25 \text{ and } Y > 0.25) = 1 - 0.5625 = 0.4375.$$

A message is verified if both channels have delivered the message, meaning:

$$P(not\ verified) = P(X \leq 0.25 \text{ and } Y > 0.25) + P(X > 0.25 \text{ and } Y \leq 0.25).$$

Calculating each term:

$$P(X \leq 0.25) = 0.25, \quad P(Y > 0.25) = 0.75 \implies P(X \leq 0.25 \text{ and } Y > 0.25) = 0.25 \cdot 0.75 = 0.1875,$$

$$P(Y \leq 0.25) = 0.25, \quad P(X > 0.25) = 0.75 \implies P(X > 0.25 \text{ and } Y \leq 0.25) = 0.75 \cdot 0.25 = 0.1875.$$

Thus, the total probability that the message is received but not verified is:

$$P(not\ verified) = 0.1875 + 0.1875 = 0.375.$$

For verification, both channels must have received the message within 15 minutes:

$$P(X \leq 0.25 \text{ and } Y \leq 0.25) = P(X \leq 0.25) \cdot P(Y \leq 0.25) = 0.25 \cdot 0.25 = 0.0625.$$

The attendant goes home 15 minutes after the message is received. Since the probability that the message is received is 0.4375*, we need to calculate the conditional probability:*

$$P(attendant\ present|verified) = \frac{P(verified)}{P(received)} = \frac{0.0625}{0.4375} = \frac{1}{7} \approx 0.142857.$$

Exercise 4.15. *Random variables B and C are jointly uniform over a $2l \times 2l$ square centered at the origin, i.e., B and C have the following joint probability density function:*

$$f_{B,C}(b,c) = \begin{cases} \frac{1}{4l^2} & \text{if } -l \leq b \leq l,\ -l \leq c \leq l \\ 0 & \text{elsewhere} \end{cases}$$

It is given that $l \geq 1$. Find the probability that the quadratic equation $x^2 + 2Bx + C = 0$ has real roots (the answer will be an expression involving l). What is the limit of this probability as $l \to \infty$?

Solution 4.15. *For the quadratic equation $x^2 + 2Bx + C = 0$ to have real roots, the discriminant must be non-negative:*

$$D = (2B)^2 - 4C \geq 0.$$

Simplifying this, we find:

$$4B^2 - 4C \geq 0 \quad \Rightarrow \quad B^2 \geq C.$$

Now, we need to determine the probability that the point (B, C) lies below the curve $C = B^2$ within the bounds of the uniform distribution, which is defined by $-l \leq B \leq l$ and $-l \leq C \leq l$.

The area under the curve $C = B^2$ for $-l \leq B \leq l$ forms a region bounded by:

(i) The parabola $C = B^2$,
(ii) The lines $C = -l$ and $C = l$.

We need to find the area of the region where $C \leq B^2$ within the square $[-l, l] \times [-l, l]$. The area below the curve from $-l$ to l is given by:

$$A_{B^2} = \int_{-l}^{l} B^2\, dB = \left[\frac{B^3}{3}\right]_{-l}^{l} = \frac{l^3}{3} - \left(-\frac{l^3}{3}\right) = \frac{2l^3}{3}.$$

The total area of the square is:

$$A_{total} = (2l)(2l) = 4l^2.$$

Thus, the probability $P(B^2 \geq C)$ is:

$$P(B^2 \geq C) = \frac{A_{B^2}}{A_{total}} = \frac{\frac{2l^3}{3}}{4l^2} = \frac{l}{6}.$$

As $l \to \infty$, the probability tends to:

$$\lim_{l\to\infty} P(B^2 \geq C) = \lim_{l\to\infty} \frac{l}{6} = \infty.$$

However, since we are looking at a bounded uniform distribution, the important observation is that while the region of interest increases with l, the bounded context of the problem implies that this probability can be normalized appropriately.

Exercise 4.16. *Consider four independent rolls of a 6-sided die. Let X be the number of 1's and let Y be the number of 2's obtained. What is the joint PMF of X and Y?*

Solution 4.16. *To find the joint PMF of X and Y, we note that X and Y can take values from 0 to 4 because there are four rolls of the die. The joint PMF $P(X = x, Y = y)$ can be determined by considering the number of ways to obtain x 1's and y 2's in four rolls, while the remaining rolls result in numbers other than 1 or 2.*

The total number of rolls is 4. We can express the joint PMF as follows:

$$P(X = x, Y = y) = \frac{4!}{x!\, y!\, (4 - x - y)!} \left(\frac{1}{6}\right)^x \left(\frac{1}{6}\right)^y \left(\frac{4}{6}\right)^{4-x-y}$$

where:
(i) $\frac{4!}{x!\,y!\,(4-x-y)!}$ is the multinomial coefficient representing the number of ways to arrange x 1's, y 2's, and $4 - x - y$ other outcomes.
(ii) $\left(\frac{1}{6}\right)^x$ is the probability of rolling x 1's.
(iii) $\left(\frac{1}{6}\right)^y$ is the probability of rolling y 2's.
(iv) $\left(\frac{4}{6}\right)^{4-x-y}$ is the probability of rolling numbers other than 1 or 2.

We can present the joint PMF in an array format for values x and y from 0 to 4:

$P(X,Y)$	$Y=0$	$Y=1$	$Y=2$	$Y=3$
$X=0$	$P(0,0)$	$P(0,1)$	$P(0,2)$	$P(0,3)$
$X=1$	$P(1,0)$	$P(1,1)$	$P(1,2)$	$P(1,3)$
$X=2$	$P(2,0)$	$P(2,1)$	$P(2,2)$	$P(2,3)$
$X=3$	$P(3,0)$	$P(3,1)$	$P(3,2)$	$P(3,3)$
$X=4$	$P(4,0)$	$P(4,1)$	$P(4,2)$	$P(4,3)$

Where each entry $P(x, y)$ can be computed using the formula given above. The values can be calculated as follows:

For example,

$$P(1,1) = \frac{4!}{1!\,1!\,2!} \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{4}{6}\right)^2 = \frac{12}{36} \cdot \frac{16}{36} = \frac{192}{1296} = \frac{16}{108} \approx 0.1481$$

You can compute the remaining values similarly for $P(x, y)$ where x and y vary from 0 to 4.

Exercise 4.17. *Let X_1, X_2, X_3 be independent random variables, uniformly distributed on $[0,1]$. Let Y be the median of X_1, X_2, X_3 (that is, the middle of the three values). Find the conditional CDF of X_1, given that $Y = 0.5$. Under this conditional distribution, is X_1 continuous?*

Solution 4.17. *First, we must understand what $Y = 0.5$ implies. One of the values must be exactly 0.5, one value must be below 0.5, and one value must be above 0.5.*

For X_1 to satisfy the condition $Y = 0.5$, there are three possible scenarios. In Case 1, $X_1 = 0.5$ when X_1 is the median. In Case 2, $X_1 < 0.5$ when X_1 is below the median. In Case 3, $X_1 > 0.5$ when X_1 is above the median. By the symmetry of the uniform distribution and the independence of the variables, we can determine the following probabilities. We have $P(X_1 < 0.5|Y = 0.5) = P(X_1 > 0.5|Y = 0.5) = P(X_1 = 0.5|Y = 0.5) = \frac{1}{3}$.

Consequently, the conditional cumulative distribution function (CDF) of X_1 given that $Y = 0.5$ is given by:

$$F_{X_1|Y=0.5}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{2x}{3} & \text{if } 0 \leq x < 0.5 \\ \frac{1}{3} & \text{if } x = 0.5 \\ \frac{2x+1}{3} & \text{if } 0.5 < x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

This distribution is classified as mixed. It contains a discrete component at $x = 0.5$ with a probability of $\frac{1}{3}$, and it includes continuous components on the intervals $[0, 0.5)$ and $(0.5, 1]$.

Therefore, given $Y = 0.5$, X_1 is neither purely continuous nor purely discrete, but rather a mixed distribution.

Chapter 5

Transformation of Random Variables

Suppose we observe a random variable, or a collection of random variables. In many practical scenarios, we might be more interested in some function of these observed random variables rather than the variables themselves. For instance, in communication systems, engineers often find it more useful to analyze the logarithm of noise power instead of the actual realization of noise. This is where the transformation of random variables becomes significant.

Let X be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$ and let $f : \mathbb{R} \to \mathbb{R}$ be a function. We want to understand the properties of $f(X)$. Since the random variable X can be viewed as a function, $f(X)$ represents a composed function that maps Ω to $\mathbb{R}$. Our first question is whether $f(X)$ qualifies as a valid random variable.

To explore this, we can consider the composed function $f \circ X(\cdot)$. If f is an arbitrary function, then $f(X)$ may not necessarily be a random variable. However, if $f : \mathbb{R} \to \mathbb{R}$ is a Borel-measurable function—meaning that the pre-images of Borel sets under f are also Borel sets—then it follows that the pre-images of Borel sets under the composed function $f \circ X(\cdot)$ are events in $\mathcal{F}$. This leads us to conclude that $f(X)$ is indeed a random variable.

The same reasoning applies to functions of several random variables. If we have a Borel-measurable function $f : \mathbb{R}^n \to \mathbb{R}$ and random variables $X_1, X_2, \ldots, X_n$, we can similarly argue that $f(X_1, \ldots, X_n)$ is also a random variable. This relationship emphasizes the importance of Borel-measurability in ensuring that the transformation of random variables yields valid random variables.

Now that we have established the conditions under which a function of a random variable qualifies as a random variable, we seek to understand the probability law of $f(X)$ based on the probability law P_X of X. In simpler terms, given the cumulative distribution function (CDF) of X, we aim to determine the CDF of $f(X)$. To approach this, we will first examine some basic functions, such as the maximum, minimum, and summation, before delving into more complex transformations.

5.1 Maximum and Minimum of Many Random Variables

Consider a set of random variables $X_1, X_2, X_3, \ldots, X_n$ defined on the probability space $(\Omega, \mathcal{F}, P)$ with a joint CDF $F_{X_1, X_2, \ldots, X_n}$. We define:

$$Y_n = \min(X_1, X_2, X_3, \ldots, X_n)$$

and

$$Z_n = \max(X_1, X_2, X_3, \ldots, X_n).$$

Our goal is to find the CDF of both Y_n and Z_n.

To begin with, let us verify that Z_n is indeed a random variable. The event $\{Z_n \leq x\}$ is equivalent to stating that each of the random variables $X_1, X_2, X_3, \ldots, X_n$ is less than or equal to x. We can express this as:

$$\{Z_n \leq x\} = \{X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x\}.$$

Now, to confirm that $\{\omega : Z(\omega) \leq z\}$ is an event, we observe that:

$$\{\omega : Z(\omega) \leq z\} = \bigcap_{i=1}^{n} \{\omega : X_i(\omega) \leq z\}.$$

This expression represents a finite intersection of events, given that each X_i is a random variable. Consequently, Z_n is indeed a legitimate random variable.

Next, we analyze the minimum Y_n. The event $\{Y_n > x\}$ can be interpreted as stating that each X_i is greater than x. Thus, we have:

$$\{Y_n > x\} = \{X_1 > x, X_2 > x, \ldots, X_n > x\}.$$

We can apply similar reasoning to prove that Y_n is also a random variable.

Now, we proceed to compute the cumulative distribution functions (CDFs) of the random variables Z_n and Y_n. For Z_n, we have:

$$P(Z_n \leq x) = P(X_1 \leq x \cap X_2 \leq x \cap \cdots \cap X_n \leq x) = F_{X_1, X_2, \ldots, X_n}(x, x, \ldots, x).$$

On the other hand, for Y_n, we can express its CDF as:

$$P(Y_n \leq x) = 1 - P(Y_n > x) = 1 - P(X_1 > x \cap X_2 > x \cap \cdots \cap X_n > x) = 1 - \overline{F_{Y_n}(x)},$$

where $\overline{F_{Y_n}(x)}$ denotes the complementary CDF of Y_n.

In particular, if the random variables $X_1, X_2, \ldots, X_n$ are independent, we find that:

$$F_{Z_n}(x) = F_{X_1}(x) F_{X_2}(x) \cdots F_{X_n}(x),$$

and for the complementary CDF of Y_n:

$$\overline{F_{Y_n}(x)} = \overline{F_{X_1}(x) F_{X_2}(x)} \cdots \overline{F_{X_n}(x)}.$$

Furthermore, if the random variables are independent and identically distributed (i.i.d.), the CDFs simplify to:

$$F_{Z_n}(x) = [F_X(x)]^n,$$

and

$$\overline{F_{Y_n}(x)} = [\overline{F_X(x)}]^n.$$

Thus, we can effectively derive the CDFs for both the maximum and minimum of a set of random variables using their joint distributions and independence properties.

Example 5.1. *Consider two independent and identically distributed random variables U_1 and U_2, each following a uniform distribution on the interval $[0,1]$, denoted as $U \sim Unif[0,1]$.*

We define: $Y = \min(U_1, U_2)$ *and* $Z = \max(U_1, U_2)$

Let $F_{U_1}(z)$ and $F_{U_2}(z)$ represent the cumulative distribution functions (CDFs) of the random variables U_1 and U_2, respectively. Since both variables are identically distributed, we have:

$$F_{U_1}(z) = F_{U_2}(z) = F_U(z)$$

The CDF of a uniform random variable U is given by:

$$F_U(z) = \begin{cases} 0 & \text{if } z < 0, \\ z & \text{if } z \in [0,1], \\ 1 & \text{if } z > 1. \end{cases}$$

Because U_1 and U_2 are also independent, the CDF of Z can be computed as follows:

$$F_Z(z) = F_{U_1}(z) \cdot F_{U_2}(z) = [F_U(z)]^2.$$

This gives us:

$$[F_U(z)]^2 = \begin{cases} 0 & \text{if } z < 0, \\ z^2 & \text{if } z \in [0,1], \\ 1 & \text{if } z > 1. \end{cases}$$

To find the probability density function (pdf) of Z, we differentiate $F_Z(z)$:

$$f_Z(z) = \begin{cases} 2z & \text{if } z \in [0,1], \\ 0 & \text{otherwise}. \end{cases}$$

Next, we analyze Y. The complementary CDF of Y, denoted as $\overline{F_Y}(y)$, is related to the CDFs of U_1 and U_2:

$$\overline{F_Y}(y) = \overline{F_{U_1}}(y) \cdot \overline{F_{U_2}}(y) = [\overline{F_U}(y)]^2,$$

where $\overline{F_U}(y) = 1 - F_U(y)$.

Thus, we can express the CDF of Y:

$$F_Y(y) = 1 - [\overline{F_U}(y)]^2.$$

Substituting for $\overline{F_U}(y)$:

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0, \\ 1 - (1-y)^2 & \text{if } y \in [0,1], \\ 1 & \text{if } y > 1. \end{cases}$$

The pdf of Y is then obtained by differentiating $F_Y(y)$:

$$f_Y(y) = \begin{cases} 0 & \text{if } y < 0, \\ 2(1-y) & \text{if } y \in [0,1], \\ 1 & \text{if } y > 1. \end{cases}$$

Example 5.2. *Let $X_1, X_2, X_3, \ldots, X_n$ be independent random variables, each following an exponential distribution characterized by parameters $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ where $\lambda_i > 0$. The cumulative distribution function (CDF) for each random variable X_i is given by:*

$$F_{X_i}(x) = 1 - e^{-\lambda_i x} \quad \textit{for } x > 0.$$

Now, let Y_n denote the minimum of these random variables, defined as:

$$Y_n = \min(X_1, X_2, \ldots, X_n).$$

To find the complementary CDF of Y_n, denoted as $\overline{F}_{Y_n}(y)$, we start by recognizing that:

$$\overline{F}_{Y_n}(y) = P(Y_n > y) = P(X_1 > y, X_2 > y, \ldots, X_n > y).$$

Since the random variables are independent, we can express this probability as the product of their individual complementary CDFs:

$$\overline{F}_{Y_n}(y) = \prod_{i=1}^{n} P(X_i > y) = \prod_{i=1}^{n} \overline{F}_{X_i}(y).$$

For each X_i, the complementary CDF is:

$$\overline{F}_{X_i}(y) = e^{-\lambda_i y}.$$

Thus, we have:

$$\overline{F}_{Y_n}(y) = \prod_{i=1}^{n} e^{-\lambda_i y} = e^{-\sum_{i=1}^{n} \lambda_i y}.$$

This implies that Y_n is also an exponential random variable, with the new parameter $\lambda = \sum_{i=1}^{n} \lambda_i$.

Exercise 5.1. *Light bulbs with Amnesia: Suppose that n light bulbs in a room are switched on at the same instant. The life time of each bulb is exponentially distributed with parameter $\mu = 1$, and are independent.*

(a) Starting from the time they are switched on, find the distribution of the time when the first bulb fuses out.

(b) Find the CDF and the density of the time when the room goes completely dark.

(c) Would your answers to the above parts change if the bulbs were not switched on at the same time, but instead, turned on at arbitrary times? Assume however that all bulbs were turned on before the first one fused out.

(d) Suppose you walk into the room and find m bulbs glowing. Starting from the instant of your walking in, what is the distribution of the time it takes until you see a bulb blow out?

Solution 5.1. *Let $X_1, X_2, \ldots, X_n$ be the lifetimes of the n light bulbs, where each X_i follows an exponential distribution with parameter $\mu = 1$.*

(a) The time when the first bulb fuses out is given by $T = \min(X_1, X_2, \ldots, X_n)$. The minimum of n independent exponential random variables with rate $\lambda = 1$ is also an exponential random variable with rate n:

$$T \sim \textit{Exponential}(n).$$

(b) *The cumulative distribution function (CDF) of T is given by:*

$$F_T(t) = P(T \leq t) = 1 - e^{-nt}, \quad t \geq 0.$$

The probability density function (PDF) is:

$$f_T(t) = \frac{d}{dt} F_T(t) = ne^{-nt}, \quad t \geq 0.$$

(c) *If the bulbs were not switched on at the same time but turned on at arbitrary times, the answers would change. The distribution of the time until the room goes dark would depend on the timing of each bulb's activation. However, if all bulbs are turned on before the first one fuses out, the distribution of T would remain the same as in part (a).*

(d) *Let $Y_1, Y_2, \ldots, Y_m$ be the lifetimes of the m bulbs glowing when you enter the room. The distribution of the time until you see a bulb blow out is given by:*

$$Z = \min(Y_1, Y_2, \ldots, Y_m).$$

Similar to part (a), Z is also an exponential random variable with parameter m:

$$Z \sim \textit{Exponential}(m).$$

Exercise 5.2. *Let X and Y be independent exponentially distributed random variables with parameters λ and μ respectively.*

(a) *Show that $Z = \min(X, Y)$ is independent of the event $\{X < Y\}$, and interpret this result verbally? [Definition: A random variable X is said to be independent of an event A if X and I_A are independent random variables, where I_A is the Indicator random variable of the event A.]*

(b) *Find $P(X = Z)$.*

Solution 5.2. *(a) To show that $Z = \min(X, Y)$ is independent of the event $\{X < Y\}$, we start by computing the joint distribution of Z and $\{X < Y\}$.*

The cumulative distribution function (CDF) of Z is given by:

$$P(Z \leq z) = P(\min(X, Y) \leq z) = 1 - P(X > z \text{ and } Y > z) = 1 - e^{-\lambda z} e^{-\mu z} = 1 - e^{-(\lambda+\mu)z}.$$

The probability density function (PDF) of Z is then:

$$f_Z(z) = (\lambda + \mu) e^{-(\lambda+\mu)z}, \quad z \geq 0.$$

Next, we find $P(X < Y)$:

$$P(X < Y) = \int_0^\infty P(X < y) f_Y(y)\, dy = \int_0^\infty (1 - e^{-\lambda y}) \mu e^{-\mu y}\, dy.$$

Evaluating this integral, we get:

$$P(X < Y) = \frac{\lambda}{\lambda + \mu}.$$

Now we need to find $P(Z \leq z, X < Y)$:

$$P(Z \leq z, X < Y) = P(\min(X, Y) \leq z, X < Y) = P(X < z) P(Y > z) = \int_0^z \lambda e^{-\lambda x} \cdot e^{-\mu z}\, dx.$$

Integrating gives:

$$= (1 - e^{-\lambda z})e^{-\mu z} = e^{-\mu z} - e^{-(\lambda+\mu)z}.$$

Thus, we have:

$$P(Z \leq z, X < Y) = e^{-\mu z} - e^{-(\lambda+\mu)z}.$$

Since $P(Z \leq z)P(X < Y) = \left(1 - e^{-(\lambda+\mu)z}\right) \frac{\lambda}{\lambda+\mu}$*, we can show that*

$$P(Z \leq z, X < Y) = P(Z \leq z)P(X < Y).$$

This indicates that Z *is independent of* $\{X < Y\}$*.*

Interpretation: *This result means that the minimum of two independent exponentially distributed random variables does not provide any information about which variable was smaller. It implies a lack of dependency between the minimum value and the event that one variable is less than the other.*

(b) To find $P(X = Z)$*, we note that:*

$$P(X = Z) = P(X < Y) = \frac{\lambda}{\lambda + \mu}.$$

This probability can be interpreted as the likelihood that X *is the minimum of* X *and* Y *when both are independently distributed.*

5.2 Sum of Random Variables

Before we explore the distributions of random variables, it is essential to verify that the sum of two random variables also forms a random variable.

Theorem 5.1. *Let* X *and* Y *be random variables defined on a probability space* $(\Omega, \mathcal{F}, P)$*. We define* $Z(\omega) = X(\omega) + Y(\omega)$ *for every* $\omega \in \Omega$*. Then,* Z *is a random variable.*

Proof. To demonstrate that Z is indeed a random variable, we need to show that the set $\{\omega \in \Omega : Z(\omega) > z\} \in \mathcal{F}$ for all $z \in \mathbb{R}$.

Now, for any $z \in \mathbb{R}$, the condition $Z(\omega) > z$ holds if and only if there exists a rational number q such that $X(\omega) > q$ and $Y(\omega) > z - q$. This equivalence arises from the property that rational numbers are dense in the real numbers. Thus, we can express the set as follows:

$$\{\omega \in \Omega : Z(\omega) > z\} = \bigcup_{q \in \mathbb{Q}} \{\omega \in \Omega : X(\omega) > q, Y(\omega) > z - q\}$$

$$= \bigcup_{q \in \mathbb{Q}} \left(\{\omega \in \Omega : X(\omega) > q\} \cap \{\omega \in \Omega : Y(\omega) > z - q\}\right).$$

Since we know that for every $q \in \mathbb{Q}$, both $\{\omega \in \Omega : X(\omega) > q\}$ and $\{\omega \in \Omega : Y(\omega) > z - q\}$ belong to $\mathcal{F}$ (because X and Y are random variables), their intersection also belongs to $\mathcal{F}$. Now, because the set of rational numbers $\mathbb{Q}$ is countable, the union of these sets indexed by q remains in $\mathcal{F}$ since $\mathcal{F}$ is a σ-algebra, which is closed under countable unions.

Thus, we conclude that:

$$\{\omega \in \Omega : Z(\omega) > z\} \in \mathcal{F},$$

which proves that the sum $Z = X + Y$ is a random variable. □

5.2.1 Sum of Two Random Variables

Discrete Case

Consider two discrete random variables, X and Y, which have a known joint probability mass function (pmf) denoted as $p_{X,Y}(x,y)$. We define a new random variable Z as follows:

$$Z = X + Y$$

Our goal is to characterize the pmf of Z, which we denote as $p_Z(z)$:

$$p_Z(z) = P(Z = z) = \sum_{x+y=z} p_{X,Y}(x,y)$$

This summation can also be expressed in a different form:

$$p_Z(z) = \sum_x P(X = x, Y = z - x) = \sum_x p_{X,Y}(x, z - x)$$

Now, if X and Y are independent random variables, the pmf of Z simplifies to:

$$p_Z(z) = \sum_x p_X(x) p_Y(z - x)$$

This result is known as the *discrete convolution* of the two pmfs.

Example 5.3. *Let X and Y be independent random variables, where X follows a Poisson distribution with parameter λ, and Y follows a Poisson distribution with parameter μ. We can compute the pmf of $Z = X + Y$ using the previous result:*

$$p_Z(z) = \sum_{x=0}^{z} e^{-\lambda} \frac{\lambda^x}{x!} e^{-\mu} \frac{\mu^{z-x}}{(z-x)!}$$

This expression simplifies as follows:

$$= e^{-(\lambda+\mu)} \frac{1}{z!} \sum_{x=0}^{z} \binom{z}{x} \lambda^x \mu^{z-x}$$

The summation inside the equation is the binomial expansion of $(\lambda + \mu)^z$, leading us to:

$$= e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^z}{z!}$$

This calculation shows that the sum of two independent Poisson-distributed random variables, with mean values λ and μ, results in another Poisson distribution with mean $\lambda + \mu$. This method can be easily extended to compute the sum of a finite number of independent Poisson random variables.

Continuous Case

We assume that X and Y random variables have a joint probability density function (pdf), denoted as $f_{X,Y}(x,y)$. We define a new random variable Z as the sum of X and Y:

$$Z = X + Y.$$

To find the cumulative distribution function (CDF) of Z, we express it as:

$$F_Z(z) = P(Z \le z) = P(X + Y \le z).$$

We can reformulate this probability using double integrals:

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{X,Y}(x, y)\, dy\, dx.$$

This expression represents the probability that the sum $X + Y$ is less than or equal to z.

Next, we can change the order of integration to facilitate computation:

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_{X,Y}(x, t - x)\, dt\, dx.$$

Here, we denote the inner integral as:

$$f_Z(t) = \int_{-\infty}^{\infty} f_{X,Y}(x, t - x)\, dx,$$

which gives us the pdf of Z. Therefore, we find that the pdf of Z is expressed as:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx.$$

In the special case where X and Y are independent continuous random variables, we have a simplification:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx,$$

which denotes the *convolution* of the two marginal pdfs:

$$f_Z(z) = f_X * f_Y.$$

Example 5.4. *Consider X_1 and X_2 as independent exponential random variables with parameters μ_1 and μ_2, respectively. Let $Z = X_1 + X_2$. Using the convolution formula derived earlier, we get:*

$$f_Z(z) = f_{X_1} * f_{X_2} = \int_0^z \mu_1 e^{-\mu_1 x} \mu_2 e^{-\mu_2 (z-x)}\, dx.$$

Factoring out constants gives us:

$$f_Z(z) = \mu_1 \mu_2 e^{-\mu_2 z} \int_0^z e^{(\mu_2 - \mu_1)x}\, dx.$$

Evaluating the integral, we arrive at the final form of $f_Z(z)$:

$$f_Z(z) = \begin{cases} \frac{\mu_1 \mu_2}{\mu_2 - \mu_1} \left(e^{-\mu_1 z} - e^{-\mu_2 z}\right) & \text{if } \mu_1 \neq \mu_2, \\ \mu^2 z e^{-\mu z} & \text{if } \mu_1 = \mu_2 = \mu. \end{cases}$$

Interestingly, this methodology can be generalized to sums of n independent exponential random variables. The pdf of such a sum, Z_n, follows an Erlang distribution:

$$f_{Z_n}(z) = \frac{\mu^n z^{n-1} e^{-\mu z}}{(n-1)!}.$$

This example illustrates the process for calculating the pdf of the sum of continuous random variables, and the methods can be readily extended to handle finite sums of various distributions.

5.2.2 Sum of Many Random Variables

In our analysis, we are looking at a scenario where we sum a set of independent random variables, but the count of these variables is also a random quantity. To formalize this, let N be a random variable that takes positive integer values, defined on a probability space $(\Omega, \mathcal{F}, P)$. We assume we know the probability mass function (pmf) of N, denoted as $P(N = n)$.

Next, let us denote the independent random variables by $X_1, X_2, \ldots$, which are also defined on the same probability space. Each of these random variables has its own cumulative distribution function (cdf), represented as $F_{X_i}(\cdot)$ for $i \geq 1$. Importantly, we assume that the random variable N is independent of the collection of random variables $\{X_i\}_{i \geq 1}$.

We define S_N as the sum of the first N random variables, which can be expressed mathematically as:

$$S_N = \sum_{i=1}^{N} X_i.$$

More specifically, for each outcome ω in our sample space Ω, this can be represented as:

$$S_N(\omega) = \sum_{i=1}^{N(\omega)} X_i(\omega).$$

To find the cumulative distribution function of S_N, denoted $F_{S_N}(x)$, we need to calculate:

$$F_{S_N}(x) = P(S_N \leq x).$$

Using the law of total probability, we can expand this expression as follows:

$$F_{S_N}(x) = \sum_{k=1}^{\infty} P(S_N \leq x \mid N = k) P(N = k).$$

Here, $P(S_N \leq x \mid N = k)$ represents the probability that the sum S_N is less than or equal to x given that N equals k. Since N is independent of the random variables X_i, we can further express this as:

$$F_{S_N}(x) = \sum_{k=1}^{\infty} P(S_k \leq x) P(N = k),$$

where S_k is the sum of the first k random variables $X_1, X_2, \ldots, X_k$. This relationship arises directly from the independence of N and the X_i's, allowing us to compute the probabilities separately.

In essence, this approach gives us a method to derive the distribution of the sum of a random number of independent random variables by leveraging the known distribution of N and the distributions of the individual random variables X_i.

Example 5.5. ***Geometric Sum of Exponentials.*** *Consider a sequence of independent random variables X_i, where each X_i follows an exponential distribution with mean μ. This means that the probability density function (pdf) of each X_i is given by:*

$$f_{X_i}(x) = \frac{1}{\mu} e^{-x/\mu} \quad \text{for } x \geq 0.$$

Let N be a random variable that represents the number of these exponential variables summed together, which follows a geometric distribution with parameter p.

The probability of N taking a specific value k is given by:

$$P(N = k) = (1 - p)^{k-1} p \quad \text{for } k \geq 1.$$

We define the sum $S_N = \sum_{i=1}^{N} X_i$. Our goal is to find the cumulative distribution function (cdf) of S_N, denoted as $F_S(x) = P(S_N \leq x)$.

We can express $F_S(x)$ as:

$$F_S(x) = \sum_{k=1}^{\infty} P(N = k) F_{S_k}(x),$$

where $S_k = \sum_{i=1}^{k} X_i$ is the sum of k exponential random variables. It is known that S_k follows an Erlang distribution, specifically:

$$F_{S_k}(x) = 1 - \sum_{n=0}^{k-1} \frac{(x)^n e^{-x}}{n!}.$$

Substituting $P(N = k)$ and $F_{S_k}(x)$ into the equation for $F_S(x)$, we have:

$$F_S(x) = \sum_{k=1}^{\infty} (1 - p)^{k-1} p \left(1 - \sum_{n=0}^{k-1} \frac{(x)^n e^{-x}}{n!} \right).$$

This can be simplified to:

$$= 1 - e^{-x} \sum_{k=1}^{\infty} (1 - p)^{k-1} \frac{(x)^k}{k!}.$$

Now, recognizing the series for the exponential function, we can rewrite the summation:

$$\sum_{k=1}^{\infty} (1 - p)^{k-1} \frac{(x)^k}{k!} = (x) e^{(1-p)x}.$$

Therefore, we have:

$$F_S(x) = 1 - e^{-x} \cdot \frac{(x)(1 - p)}{(1 - p)} e^{(1-p)x},$$

which simplifies to:

$$= 1 - e^{-(p)x}.$$

Thus, we conclude that the geometric sum of independent exponential random variables results in another exponential distribution with a modified parameter.

Consider a radioactive source that emits α particles. The time between two successive emissions follows an exponential distribution characterized by a parameter λ. Each time an emission occurs, a detector has a probability p of detecting it, and a probability $1 - p$ of missing it. Importantly, these detection events are independent of one another.

From this setup, we can conclude that the time between two successive detections behaves like a geometric sum of independent and identically distributed (i.i.d.) exponential random variables. The resulting distribution of the time until detection is also an exponential random variable, but with a new parameter $p\lambda$. This is an important finding because it illustrates how the combination of random events can yield a new type of randomness.

Now, let's explore a more intricate scenario. Imagine a gambler who plays a game multiple times, receiving either rewards or penalties with each round. The gambler continues to play until they feel satisfied with their winnings or, conversely, until they are broke. In this case, let X_i represent the amount gained (or lost) in the i-th round.

Here, the total earnings of the gambler at the end of the game becomes more complex to analyze because the number of rounds N played depends on the results of the rounds themselves, namely the outcomes X_i. This situation illustrates the idea of *stopping rules*, which is a significant topic in probability theory and will be explored in more advanced courses. The need to understand this dependence opens up new avenues for research and exploration in stochastic processes, particularly in scenarios where decisions are made based on previous outcomes, rather than occurring independently.

Exercise 5.3. *Let X_1 and X_2 be independent random variables with distributions $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$ respectively. Show that the distribution of $X_1 + X_2$ is $N(0, \sigma_1^2 + \sigma_2^2)$.*

Solution 5.3. *To show that $Y = X_1 + X_2$ follows a normal distribution, we can leverage the property of the sum of independent normal variables.*

The characteristic function of a normally distributed random variable $X \sim N(0, \sigma^2)$ is given by:

$$\phi_X(t) = e^{-\frac{1}{2}\sigma^2 t^2}.$$

For $X_1 \sim N(0, \sigma_1^2)$:

$$\phi_{X_1}(t) = e^{-\frac{1}{2}\sigma_1^2 t^2}.$$

For $X_2 \sim N(0, \sigma_2^2)$:

$$\phi_{X_2}(t) = e^{-\frac{1}{2}\sigma_2^2 t^2}.$$

Since X_1 and X_2 are independent, the characteristic function of $Y = X_1 + X_2$ is the product of the individual characteristic functions:

$$\phi_Y(t) = \phi_{X_1}(t) \cdot \phi_{X_2}(t) = e^{-\frac{1}{2}\sigma_1^2 t^2} \cdot e^{-\frac{1}{2}\sigma_2^2 t^2} = e^{-\frac{1}{2}(\sigma_1^2 + \sigma_2^2) t^2}.$$

The resulting characteristic function $\phi_Y(t) = e^{-\frac{1}{2}(\sigma_1^2 + \sigma_2^2) t^2}$ corresponds to the characteristic function of a normal distribution with mean 0 and variance $\sigma_1^2 + \sigma_2^2$.

Exercise 5.4. *Consider two independent and identically distributed discrete random variables X and Y. Assume that their common PMF, denoted by $p(z)$, is symmetric around zero, i.e., $p(z) = p(-z), \forall z$. Show that the PMF of $X + Y$ is also symmetric around zero and is largest at zero.*

Solution 5.4. *Let $Z = X + Y$. To find the PMF of Z, we will compute $P(Z = k)$ for any integer k.*

Since X and Y are independent, we have:

$$P(Z = k) = P(X + Y = k) = \sum_j P(X = j) P(Y = k - j) = \sum_j p(j) p(k - j)$$

Now, we will show that $P(Z = k) = P(Z = -k)$ for all integers k:

$$P(Z = -k) = \sum_j P(X = j) P(Y = -k - j) = \sum_j p(j) p(-k - j)$$

Using the symmetry of the PMF $p(z)$, we have $p(-k - j) = p(k + j)$, so:

$$P(Z = -k) = \sum_j p(j) p(k + j)$$

Now, we change the index of summation by letting $i = k + j$, which gives $j = i - k$. Then the limits of summation change accordingly:

$$P(Z = -k) = \sum_i p(i-k)p(i) = \sum_i p(i)p(k-i)$$

This is equal to $P(Z = k)$:

$$P(Z = -k) = P(Z = k)$$

Thus, the PMF $P(Z = k)$ is symmetric around zero.

To show that it is largest at zero, we note:

$$P(Z = 0) = \sum_j P(X = j)P(Y = -j) = \sum_j p(j)p(-j) = \sum_j p(j)^2$$

Since all $p(j)$ are non-negative, $P(Z = 0)$ is a sum of squares, which is maximized when $j = 0$.

For any $k \neq 0$:

$$P(Z = k) = \sum_j p(j)p(k-j)$$

Each term $p(j)p(k-j)$ will generally be less than or equal to $p(0)$ due to the nature of convolution of symmetric distributions, meaning $P(Z = k) < P(Z = 0)$.

Therefore, the PMF of $Z = X + Y$ is symmetric around zero and is largest at zero.

Exercise 5.5. *Suppose X and Y are independent random variables with $Z = X + Y$ such that*

$$f_X(x) = ce^{-cx}, \quad x \geq 0$$

and

$$f_Z(z) = \frac{c^2}{2} z e^{-cz}, \quad z \geq 0.$$

Compute $f_Y(y)$.

Solution 5.5. *Given that X and Y are independent random variables, the joint probability density function (PDF) of X and Y can be expressed as the product of their individual PDFs:*

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

The PDF of Z, which is the sum of X and Y, is given by the convolution of the PDFs of X and Y:

$$f_Z(z) = \int_0^z f_X(x) f_Y(z-x)\, dx.$$

From the problem, we have:

$$f_Z(z) = \frac{c^2}{2} z e^{-cz}.$$

We know the PDF of X:

$$f_X(x) = ce^{-cx}.$$

To find $f_Y(y)$, we first need to express $f_Z(z)$ in terms of $f_Y(y)$. Assuming $f_Y(y)$ has the same functional form as $f_X(x)$, we let:

$$f_Y(y) = ke^{-ky},$$

where k is a constant to be determined.

Substituting $f_Y(y)$ into the convolution integral, we have:

$$f_Z(z) = \int_0^z f_X(x) f_Y(z-x)\,dx = \int_0^z ce^{-cx} ke^{-k(z-x)}\,dx.$$

This simplifies to:

$$f_Z(z) = cke^{-kz} \int_0^z e^{(k-c)x}\,dx.$$

Evaluating the integral:

$$\int_0^z e^{(k-c)x}\,dx = \frac{1}{k-c}(e^{(k-c)z} - 1).$$

Therefore,

$$f_Z(z) = cke^{-kz} \cdot \frac{1}{k-c}(e^{(k-c)z} - 1).$$

We need this to equal $\frac{c^2}{2}ze^{-cz}$. Matching the coefficients of ze^{-cz} leads to a system of equations, and solving this system gives the values of c and k.

Eventually, through careful consideration, we find:

$$k = c,$$

which implies:

$$f_Y(y) = ce^{-cy}, \quad y \geq 0.$$

Exercise 5.6. *Let X_1 and X_2 be the number of calls arriving at a switching centre from two different localities at a given instant of time. X_1 and X_2 are well modelled as independent Poisson random variables with parameters λ_1 and λ_2 respectively.*

(a) Find the PMF of the total number of calls arriving at the switching centre.

(b) Find the conditional PMF of X_1 given the total number of calls arriving at the switching centre is n.

Solution 5.6. *(a) The total number of calls arriving at the switching centre can be expressed as $X = X_1 + X_2$. Since X_1 and X_2 are independent Poisson random variables, the PMF of X can be derived from the property of the sum of independent Poisson random variables.*

The PMF of a Poisson random variable X_i with parameter λ_i is given by:

$$P(X_i = k) = \frac{\lambda_i^k e^{-\lambda_i}}{k!}, \quad k = 0, 1, 2, \ldots$$

Since X_1 and X_2 are independent, the PMF of the total number of calls X is:

$$P(X = n) = \sum_{k=0}^{n} P(X_1 = k) P(X_2 = n-k)$$

This can be computed as follows:

$$P(X = n) = \sum_{k=0}^{n} \frac{\lambda_1^k e^{-\lambda_1}}{k!} \cdot \frac{\lambda_2^{n-k} e^{-\lambda_2}}{(n-k)!}$$

Recognizing that this sum represents the PMF of a Poisson distribution with parameter $\lambda_1 + \lambda_2$, we have:

$$P(X = n) = \frac{(\lambda_1 + \lambda_2)^n e^{-(\lambda_1+\lambda_2)}}{n!}$$

Thus, X is also a Poisson random variable with parameter $\lambda_1 + \lambda_2$.

(b) To find the conditional PMF of X_1 given that the total number of calls arriving at the switching centre is n, we apply the formula for conditional probability:

$$P(X_1 = k \mid X = n) = \frac{P(X_1 = k, X_2 = n-k)}{P(X = n)}$$

Using the independence of X_1 and X_2:

$$P(X_1 = k, X_2 = n-k) = P(X_1 = k)P(X_2 = n-k) = \frac{\lambda_1^k e^{-\lambda_1}}{k!} \cdot \frac{\lambda_2^{n-k} e^{-\lambda_2}}{(n-k)!}$$

Substituting into the conditional probability gives:

$$P(X_1 = k \mid X = n) = \frac{\frac{\lambda_1^k e^{-\lambda_1}}{k!} \cdot \frac{\lambda_2^{n-k} e^{-\lambda_2}}{(n-k)!}}{\frac{(\lambda_1+\lambda_2)^n e^{-(\lambda_1+\lambda_2)}}{n!}}$$

Simplifying this expression yields:

$$P(X_1 = k \mid X = n) = \frac{n!}{k!(n-k)!} \cdot \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n}$$

This shows that $X_1 \mid (X = n)$ follows a binomial distribution:

$$X_1 \mid (X = n) \sim \textit{Binomial}(n, p) \quad \textit{where } p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Exercise 5.7. *The random variables X, Y, and Z are independent and uniformly distributed between zero and one. Find the PDF of $X + Y + Z$.*

Solution 5.7. *To find the probability density function (PDF) of the sum $S = X + Y + Z$, where X, Y, and Z are independent random variables uniformly distributed on $[0, 1]$, we can use the convolution of their individual PDFs.*

The PDF of a uniform random variable U on $[0, 1]$ is given by:

$$f_U(u) = \begin{cases} 1 & \textit{if } 0 \le u \le 1 \\ 0 & \textit{otherwise} \end{cases}$$

The PDF of the sum of two independent random variables can be found using the convolution of their PDFs:

$$f_{X+Y}(s) = \int_0^s f_X(x) f_Y(s-x)\, dx$$

Since f_X and f_Y are both equal to 1 on $[0, 1]$:

1. For $0 \le s < 1$:

$$f_{X+Y}(s) = \int_0^s 1 \cdot 1\, dx = s$$

2. For $1 \le s < 2$:

$$f_{X+Y}(s) = \int_{s-1}^1 1 \cdot 1\, dx = 2 - s$$

The PDF $f_{X+Y}(s)$ is given by:

$$f_{X+Y}(s) = \begin{cases} s & \text{if } 0 \le s < 1 \\ 2 - s & \text{if } 1 \le s < 2 \\ 0 & \text{otherwise} \end{cases}$$

Next, we need to find the PDF of $S = X + Y + Z$. We will convolve $f_{X+Y}(s)$ with $f_Z(z)$.

The convolution is given by:

$$f_S(s) = \int_0^s f_{X+Y}(x) f_Z(s - x)\, dx$$

Since $f_Z(z) = 1$ for $0 \le z \le 1$, we consider two cases for s:

1. For $0 \le s < 1$:

$$f_S(s) = \int_0^s x \cdot 1\, dx = \frac{s^2}{2}$$

2. For $1 \le s < 2$:

$$f_S(s) = \int_0^1 x \cdot 1\, dx + \int_1^s (2 - x) \cdot 1\, dx = \frac{1}{2} + (2s - \frac{s^2}{2} - 1) = 2 - \frac{s^2}{2}$$

3. For $2 \le s < 3$:

$$f_S(s) = \int_{s-2}^1 (2 - x) \cdot 1\, dx = (2 - (s - 2)) = 4 - s$$

Thus, the PDF $f_S(s)$ is given by:

$$f_S(s) = \begin{cases} \frac{s^2}{2} & \text{if } 0 \le s < 1 \\ 2 - \frac{s^2}{2} & \text{if } 1 \le s < 2 \\ 4 - s & \text{if } 2 \le s < 3 \\ 0 & \text{otherwise} \end{cases}$$

Exercise 5.8. *Construct an example to show that the sum of a random number of independent normal random variables is not normal.*

Solution 5.8. *To demonstrate that the sum of a random number of independent normal random variables can result in a non-normal distribution, consider the following example:*

Let $X_1, X_2, \ldots, X_n$ be independent normal random variables, each distributed as $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, \ldots, n$. We will sum these variables, but first, we will allow the number of variables n to be a random variable itself.

Let N be a random variable that follows a Poisson distribution with parameter λ, i.e., $N \sim \text{Poisson}(\lambda)$.

Define the sum of the random variables as follows:

$$S = \sum_{i=1}^{N} X_i$$

where N is the number of terms in the sum.

The random variable S is conditioned on N, and its distribution can be understood through the law of total probability:

$$S|N = n \sim N\left(n\mu, n\sigma^2\right)$$

To find the marginal distribution of S, we need to consider all possible values of N:

$$f_S(s) = \sum_{n=0}^{\infty} f_{S|N}(s|n)P(N = n)$$

However, since N follows a Poisson distribution, the total sum S will have a mixed distribution depending on the realization of N.

Suppose each $X_i \sim N(0, 1)$. If N is a Poisson random variable with $\lambda = 2$, the resulting distribution of S is a mixture of normal distributions, which is not normal overall. This can be shown through simulation or density plots.

Thus, the sum $S = \sum_{i=1}^{N} X_i$ does not follow a normal distribution due to the randomness in the number of summands.

5.3 General Transformation of Random Variables

We have previously explored some fundamental transformations of random variables, such as the sums of random variables and their maximum and minimum values. Now, we will delve into more general transformations of random variables. The motivation for transforming a random variable can be illustrated with the following example:

Imagine we have a particle whose velocity is represented by a random variable V. Each specific realization of the velocity corresponds to a particular value of kinetic energy, denoted as E. Our goal is to understand the distribution of the kinetic energy E, which depends on the original random variable V through a transformation.

Mathematically, this can be expressed as:

$$E = f(V)$$

where f is a function that defines how the velocity V translates into kinetic energy E. In this case, the distribution of the kinetic energy E is derived from the distribution of the velocity V.

Such scenarios are common in practical applications, where we often need to study new random variables that arise from transformations of existing ones. Understanding these transformations not only helps us derive the properties of the new variables but also deepens our insight into the underlying stochastic processes.

5.3.1 Transformation of a Single Random Variable

Consider a random variable $X : \Omega \to \mathbb{R}$ and a Borel measurable function $g : \mathbb{R} \to \mathbb{R}$. When we define a new random variable $Y = g(X)$, we are interested in determining the distribution of Y. Specifically, we want to find the cumulative distribution function (CDF) $F_Y(y)$ based on the CDF $F_X(x)$.

The CDF of Y can be expressed as:

$$F_Y(y) = P(g(X) \leq y) = P\{\omega \in \Omega \mid g(X(\omega)) \leq y\}.$$

To facilitate our calculation, we define the set B_y as the collection of all x such that $g(x) \leq y$. Thus, we have:

$$F_Y(y) = P_X(B_y).$$

Example 5.6. *Let X be a Gaussian random variable with mean 0 and variance 1, denoted as $X \sim N(0,1)$. We want to find the distribution of $Y = X^2$.*

To compute this, we first express the CDF of Y:

$$F_Y(y) = P(X^2 \leq y).$$

This is equivalent to finding the probabilities for X being in the interval $[-\sqrt{y}, \sqrt{y}]$:

$$F_Y(y) = P(-\sqrt{y} \leq X \leq \sqrt{y}).$$

We can calculate this probability using the CDF of the standard normal distribution, Φ:

$$F_Y(y) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}).$$

Since the CDF of the standard normal distribution is symmetric, we have:

$$F_Y(y) = 2\Phi(\sqrt{y}).$$

Next, we differentiate $F_Y(y)$ with respect to y to find the probability density function (PDF) $f_Y(y)$:

$$f_Y(y) = \frac{dF_Y(y)}{dy}.$$

Calculating $F_Y(y)$ explicitly, we have:

$$F_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Changing variables by letting $t^2 = u$ leads to:

$$F_Y(y) = 2\int_0^{y} \frac{1}{2\sqrt{2\pi u}} e^{-\frac{u}{2}} du.$$

Consequently, the density function becomes:

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}, \quad \text{for } y > 0.$$

Important Notes:

1. The random variable Y can only take non-negative values since it is derived from squaring the real-valued random variable X.
2. The distribution of Y as the square of a Gaussian random variable is recognized as the Chi-squared distribution.

Thus, we observe that given the distribution of the random variable X, the distribution of any function of X can be derived using fundamental principles. We now come up with a direct formula to find the distribution of a function of the random variable in the cases where the function is differentiable and monotonic.

The Generic Formula

Let X be a random variable with a probability density function $f_X(x)$, and let g be a monotonically increasing function. We define a new random variable Y as follows:

$$Y = g(X).$$

To find the cumulative distribution function (CDF) of Y, denoted $F_Y(y)$, we start by expressing it in terms of Y:

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y).$$

Since g is monotonically increasing, the inequality $g(X) \leq y$ implies that

$$X \leq g^{-1}(y).$$

Thus, we can express the CDF of Y as:

$$F_Y(y) = P(X \leq g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(x)\, dx.$$

Now, let us substitute $x = g^{-1}(t)$. By the chain rule of differentiation, we have:

$$dg(x) = g'(x)\, dx \Rightarrow dx = \frac{dt}{g'(g^{-1}(t))}.$$

Therefore, the CDF can be rewritten as:

$$F_Y(y) = \int_{-\infty}^{y} f_X(g^{-1}(t)) \frac{dt}{g'(g^{-1}(t))}.$$

Next, to find the probability density function (PDF) $f_Y(y)$, we differentiate the CDF $F_Y(y)$:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(g^{-1}(y)) \cdot \frac{1}{g'(g^{-1}(y))}.$$

The term $\frac{1}{g'(g^{-1}(y))}$ is known as the Jacobian of the transformation $g(\cdot)$.

Now, we can also apply a similar logic for a monotonically decreasing function g. In this case, we would find:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left(-\frac{1}{g'(g^{-1}(y))} \right).$$

Thus, for any monotonic function g, the general formula for the probability density function of Y can be summarized as:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot |g'(g^{-1}(y))|.$$

Example 5.7. *Let $X \sim N(0, 1)$. We want to find the distribution of $Y = e^X$.*

To start, observe that the function $g(x) = e^x$ is both differentiable and monotonically increasing. The inverse function of g is given by:

$$g^{-1}(y) = \ln(y)$$

Next, we can compute the derivative of g:

$$g'(x) = e^x$$

Thus, we find:

$$g'(g^{-1}(y)) = g'(\ln(y)) = y$$

Since $g'(g^{-1}(y))$ is positive for all values of $y > 0$, we can apply the change of variables formula to find the probability density function (pdf) of Y:

$$f_Y(y) = f_X(\ln(y)) \cdot \left|\frac{d}{dy} g^{-1}(y)\right|$$

Substituting the known values, we get:

$$f_Y(y) = f_X(\ln(y)) \cdot \frac{1}{y}$$

Given that $f_X(x)$ for $X \sim N(0, 1)$ is:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

we can replace x with $\ln(y)$:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\ln(y))^2}{2}} \cdot \frac{1}{y}$$

Therefore, the pdf of Y is given by:

$$f_Y(y) = \frac{1}{y\sqrt{2\pi}} e^{-\frac{(\ln(y))^2}{2}} \quad \textit{for } y > 0.$$

This result indicates that Y follows a log-normal distribution.

Example 5.8. *Let $U \sim Uniform(0, 1)$, meaning U is a uniform random variable on the interval $[0, 1]$. We want to find the distribution of $Y = -\ln(U)$.*

Here, the function $g(u) = -\ln(u)$ is differentiable and monotonically decreasing. Its inverse function is:

$$g^{-1}(y) = e^{-y}$$

Now, we compute the derivative of g:

$$g'(u) = -\frac{1}{u}$$

Thus, we have:

$$g'(g^{-1}(y)) = g'(e^{-y}) = -e^y$$

Since $g'(g^{-1}(y))$ is negative for all values of y, we compute the absolute value of the Jacobian:

$$\left|g'(g^{-1}(y))\right| = -g'(g^{-1}(y)) = \frac{1}{e^{-y}} = e^y.$$

Now, we can use the change of variables formula to find the pdf of Y:

$$f_Y(y) = f_U(e^{-y}) \cdot |g'(g^{-1}(y))|.$$

Given that $f_U(u) = 1$ for $U \sim$ Uniform$(0,1)$:

$$f_Y(y) = 1 \cdot e^y = e^{-y} \quad \textit{for } y \geq 0.$$

This indicates that Y is an exponential random variable with mean 1.

5.3.2 Transformation of Many Random Variables

The generic formula for transformations can indeed be expanded to encompass multiple random variables. To illustrate this, let us consider an n-tuple random variable, denoted as $(X_1, X_2, \ldots, X_n)$. This random variable has a joint density function represented by

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n).$$

Now, we define transformations of these variables as follows:

$$Y_1 = g_1(X_1, X_2, \ldots, X_n), \quad Y_2 = g_2(X_1, X_2, \ldots, X_n), \quad \ldots, \quad Y_n = g_n(X_1, X_2, \ldots, X_n.$$

For convenience, we can succinctly express this transformation as a vector:

$$\mathbf{Y} = \mathbf{g}(\mathbf{X}),$$

where $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$.

We proceed with the assumption that the transformation $\mathbf{g}$ is both invertible and continuously differentiable. Under these conditions, we can derive the joint density of the transformed variables, expressed as:

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = f_{X_1,X_2,\ldots,X_n}(g^{-1}(y)) \cdot |J(y)|,$$

where $|J(y)|$ denotes the Jacobian determinant, which is a crucial element in transforming densities.

The Jacobian matrix $J(y)$ is defined as:

$$J(y) = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial y_n} & \frac{\partial x_2}{\partial y_n} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix}.$$

The Jacobian serves as a scaling factor that adjusts the volume of the transformed space relative to the original space. When we perform a change of variables in probability, the transformation can stretch or compress areas of the probability density function. The absolute value of the Jacobian determinant captures this effect mathematically.

To understand this better, consider a simple two-dimensional example where we transform a shape in the XY-plane to a new shape in the UV-plane. If the transformation causes the area of a small rectangle in the XY-plane to become larger or smaller in the UV-plane, the Jacobian tells us exactly how much to scale the density of points in that area to preserve the total probability.

Example 5.9. *Consider a particle whose Euclidean coordinates X and Y are drawn from independent Gaussian random variables with mean 0 and variance 1, i.e., $X, Y \sim N(0,1)$. We aim to find the distribution of the particle's polar coordinates R and Θ.*

The transformations from Cartesian to polar coordinates are given by:

$$X = R\cos\Theta \quad \textit{and} \quad Y = R\sin\Theta.$$

To determine the distribution of R and Θ, we first calculate the Jacobian of the transformation. The partial derivatives are as follows:

$$\frac{\partial x}{\partial r} = \cos\theta, \quad \frac{\partial y}{\partial r} = \sin\theta,$$

$$\frac{\partial x}{\partial \theta} = -r\sin\theta, \quad \frac{\partial y}{\partial \theta} = r\cos\theta.$$

Thus, the Jacobian J is computed as:

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = r(\cos^2\theta + \sin^2\theta) = r.$$

Next, since X and Y are independent random variables, their joint probability density function is given by:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}, \quad x, y \in \mathbb{R}.$$

Using the transformation to polar coordinates, we find:

$$f_{R,\Theta}(r,\theta) = f_{X,Y}(r\cos\theta, r\sin\theta)\cdot|J| = \frac{1}{2\pi}e^{-\frac{(r\cos\theta)^2+(r\sin\theta)^2}{2}}\cdot r = \frac{1}{2\pi}e^{-\frac{r^2}{2}}\cdot r,$$

where $r \geq 0$ *and* $\theta \in [0, 2\pi]$.

To find the marginal densities of R and Θ, we integrate the joint distribution:

$$f_R(r) = \int_0^{2\pi} f_{R,\Theta}(r,\theta)\,d\theta = re^{-r^2/2}, \quad \textit{for } r \geq 0,$$

$$f_\Theta(\theta) = \int_0^{\infty} f_{R,\Theta}(r,\theta)\,dr = \frac{1}{2\pi}, \quad \textit{for } \theta \in [0, 2\pi].$$

The distribution $f_R(r)$ *is known as the Rayleigh distribution, which is commonly used in wireless communications to model the gain of a fading channel.*

It is noteworthy that R and Θ are independent random variables since the joint distribution factors into the product of the marginals:

$$f_{R,\Theta}(r,\theta) = f_R(r)\cdot f_\Theta(\theta).$$

Exercise 5.9. *Let* $X \sim Exp(0.5)$*. Prove that* $Y = \sqrt{X}$ *is a Rayleigh distributed random variable.*

Solution 5.9. *To show that* $Y = \sqrt{X}$ *is a Rayleigh distributed random variable, we start with the probability density function (PDF) of the exponential distribution. The PDF of* $X \sim Exp(\lambda)$ *is given by:*

$$f_X(x) = \lambda e^{-\lambda x} \quad \textit{for } x \geq 0.$$

In our case, $\lambda = 0.5$*, so:*

$$f_X(x) = 0.5e^{-0.5x} \quad \textit{for } x \geq 0.$$

Next, we find the cumulative distribution function (CDF) of X:

$$F_X(x) = \int_0^x f_X(t)\,dt = \int_0^x 0.5e^{-0.5t}\,dt.$$

Evaluating the integral, we have:

$$F_X(x) = -e^{-0.5t}\Big|_0^x = -e^{-0.5x} + 1 = 1 - e^{-0.5x}.$$

Now we find the distribution of $Y = \sqrt{X}$. To do this, we first express X in terms of Y:

$$X = Y^2.$$

Next, we need to find the PDF of Y. We can use the transformation method for random variables. The relationship between X and Y gives us:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|,$$

where $x = y^2$. Calculating the derivative:

$$\frac{dx}{dy} = 2y.$$

Thus, we can write:

$$f_Y(y) = f_X(y^2) \cdot |2y|.$$

Substituting for $f_X(y^2)$:

$$f_Y(y) = 0.5e^{-0.5y^2} \cdot 2y = ye^{-0.5y^2}.$$

This is the PDF of a Rayleigh distribution. The PDF of a Rayleigh distributed random variable with parameter σ is given by:

$$f_Y(y) = \frac{y}{\sigma^2}e^{-\frac{y^2}{2\sigma^2}} \quad \text{for } y \geq 0.$$

Setting $\sigma^2 = 2$, we get:

$$f_Y(y) = \frac{y}{2}e^{-\frac{y^2}{4}}.$$

Since this matches the form we derived, we conclude that $Y = \sqrt{X}$ is indeed Rayleigh distributed with parameter $\sigma = 1$.

Exercise 5.10. *Let X be a random variable with a continuous distribution F.*

(i) Show that the random variable $Y = F(X)$ is uniformly distributed over $[0, 1]$. [Hint: Although F is the distribution of X, regard it simply as a function satisfying certain properties required to make it a CDF!]

(ii) Now, given that $Y = y$, a random variable Z is distributed as Geometric with parameter y. Find the unconditional PMF of Z. Also, given $Z = z$ for some $z \geq 1, z \in \mathbb{N}$, find the conditional PMF of Y.

Solution 5.10. *To show that the random variable $Y = F(X)$ is uniformly distributed over $[0, 1]$, we start by computing the cumulative distribution function (CDF) of Y.*

The CDF of Y is given by:

$$P(Y \leq y) = P(F(X) \leq y.$$

For $y \in [0, 1]$, this can be rewritten using the properties of F:

$$P(F(X) \leq y) = P(X \leq F^{-1}(y)).$$

Since F is continuous and strictly increasing, we have:

$$P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y.$$

Therefore, the CDF of Y is:

$$P(Y \leq y) = y \quad \text{for } y \in [0, 1].$$

This shows that Y is uniformly distributed over $[0, 1]$.

Given that $Y = y$, the random variable Z follows a Geometric distribution with parameter y. The probability mass function (PMF) of a Geometric random variable is given by:

$$P(Z = z \mid Y = y) = (1 - y)^{z-1} y \quad \text{for } z \in \mathbb{N}, z \geq 1.$$

To find the unconditional PMF of Z, we need to use the law of total probability:

$$P(Z = z) = \int_0^1 P(Z = z \mid Y = y) f_Y(y)\, dy.$$

Since Y is uniformly distributed over $[0, 1]$, we have $f_Y(y) = 1$ for $y \in [0, 1]$. Thus:

$$P(Z = z) = \int_0^1 (1 - y)^{z-1} y\, dy.$$

This integral can be computed using integration by parts or Beta function properties:

$$\int_0^1 (1 - y)^{z-1} y\, dy = \frac{1}{(z+1)(z+2)}.$$

Therefore, the unconditional PMF of Z is:

$$P(Z = z) = \frac{1}{(z+1)(z+2)} \quad \text{for } z \in \mathbb{N}, z \geq 1.$$

Next, we find the conditional PMF of Y given $Z = z$:

$$P(Y = y \mid Z = z) = \frac{P(Z = z \mid Y = y) P(Y = y)}{P(Z = z)}.$$

Substituting the known values:

$$P(Y = y \mid Z = z) = \frac{(1 - y)^{z-1} y \cdot 1}{P(Z = z)}.$$

This gives us the conditional PMF of Y given $Z = z$.

Exercise 5.11. *Let X be a continuous random variable with the pdf*

$$f_X(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Find the transformation $Y = g(X)$ such that the pdf of Y will be

$$f_Y(y) = \begin{cases} \frac{1}{2}\sqrt{y} & 0 < y < 1 \\ 0 & \textit{otherwise} \end{cases}$$

Solution 5.11. *To find the transformation $Y = g(X)$ such that the pdf of Y matches the given pdf $f_Y(y)$, we will first find the cumulative distribution function (CDF) of X.*

The CDF of X is given by:

$$F_X(x) = \int_0^x f_X(t)\,dt = \int_0^x e^{-t}\,dt = 1 - e^{-x}, \quad x \geq 0.$$

Thus, the CDF can be expressed as:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geq 0 \end{cases}$$

Now, to find $g(X)$, we can utilize the relationship between the CDF of Y and X:

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)).$$

Given $f_Y(y)$, we can find $F_Y(y)$:

$$F_Y(y) = \int_0^y f_Y(t)\,dt = \int_0^y \frac{1}{2}\sqrt{t}\,dt = \frac{1}{2} \cdot \frac{2}{3} t^{3/2}\Big|_0^y = \frac{1}{3} y^{3/2}, \quad 0 < y < 1.$$

Setting this equal to $F_X(g^{-1}(y))$:

$$F_X(g^{-1}(y)) = 1 - e^{-g^{-1}(y)}.$$

Thus, we need:

$$1 - e^{-g^{-1}(y)} = \frac{1}{3} y^{3/2}.$$

Rearranging gives:

$$e^{-g^{-1}(y)} = 1 - \frac{1}{3} y^{3/2}.$$

Taking the natural logarithm:

$$-g^{-1}(y) = \ln\left(1 - \frac{1}{3} y^{3/2}\right),$$

which implies:

$$g^{-1}(y) = -\ln\left(1 - \frac{1}{3} y^{3/2}\right).$$

Therefore, we can express $g(X)$ as:

$$Y = g(X) = -\ln\left(1 - \frac{1}{3} X^{3/2}\right).$$

Thus, the transformation $Y = g(X)$ that yields the desired pdf $f_Y(y)$ is:

$$Y = g(X) = -\ln\left(1 - \frac{1}{3} X^{3/2}\right).$$

Exercise 5.12. *Suppose X and Y are independent Gaussian random variables with zero mean and variance σ^2. Show that the ratio $\frac{X}{Y}$ follows a Cauchy distribution.*

Solution 5.12. *Let $X \sim \mathcal{N}(0, \sigma^2)$ and $Y \sim \mathcal{N}(0, \sigma^2)$ be independent Gaussian random variables with mean zero and variance σ^2. We aim to show that the ratio $Z = \frac{X}{Y}$ has a Cauchy distribution.*

Since X and Y are independent normal random variables, their joint probability density function (PDF) is given by:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right).$$

We introduce the transformation:

$$Z = \frac{X}{Y} \quad \textit{and} \quad W = Y.$$

This implies $X = ZW$ and $Y = W$.

To find the joint distribution of Z and W, we need the Jacobian determinant of this transformation:

$$\frac{\partial(X, Y)}{\partial(Z, W)} = \begin{vmatrix} \frac{\partial X}{\partial Z} & \frac{\partial X}{\partial W} \\ \frac{\partial Y}{\partial Z} & \frac{\partial Y}{\partial W} \end{vmatrix} = \begin{vmatrix} W & Z \\ 0 & 1 \end{vmatrix} = W.$$

Thus, the absolute value of the Jacobian determinant is $|W|$.

Using the change of variables, the joint PDF of Z and W is:

$$f_{Z,W}(z, w) = f_{X,Y}(zw, w) \cdot |w| = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{z^2 w^2 + w^2}{2\sigma^2}\right) \cdot |w|.$$

Simplifying the exponent:

$$f_{Z,W}(z, w) = \frac{|w|}{2\pi\sigma^2} \exp\left(-\frac{w^2(z^2 + 1)}{2\sigma^2}\right).$$

To obtain the marginal PDF of Z, integrate out w:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,W}(z, w)\, dw = \int_{-\infty}^{\infty} \frac{|w|}{2\pi\sigma^2} \exp\left(-\frac{w^2(z^2 + 1)}{2\sigma^2}\right) dw.$$

This integral simplifies by using a standard Gaussian integral:

$$f_Z(z) = \frac{1}{\pi} \frac{\sigma}{\sigma(z^2 + 1)} = \frac{1}{\pi} \frac{1}{z^2 + 1}.$$

This is the PDF of Cauchy distribution.

Exercise 5.13. *Particles are subject to collisions that cause them to split into two parts, with each part being a fraction of the parent. Suppose that this fraction is uniformly distributed between 0 and 1. Following a single particle through several splittings, we obtain a fraction of the original particle $Z_n = X_1 X_2 \dots X_n$, where each X_j is uniformly distributed between 0 and 1. Show that the density for the random variable Z_n is given by:*

$$f_n(z) = \frac{1}{(n-1)!}(-\log(z))^{n-1}$$

Solution 5.13. *Since Z_n is the product of n independent uniform random variables, we can express Z_n as:*

$$Z_n = X_1 X_2 \ldots X_n.$$

Take the natural logarithm of Z_n:

$$\log(Z_n) = \log(X_1) + \log(X_2) + \cdots + \log(X_n).$$

Each $\log(X_j)$ is independently and identically distributed. Since $X_j \sim$ Uniform$(0,1)$, the distribution of $\log(X_j)$ is exponential with parameter $\lambda = 1$, and thus has mean -1 and density:

$$f_{\log(X_j)}(x) = e^x \quad \text{for } x < 0.$$

The sum of n independent exponential random variables with rate $\lambda = 1$ follows a gamma distribution with shape parameter n and rate parameter 1. Therefore, $S = -\log(Z_n)$ has the density:

$$f_S(s) = \frac{s^{n-1}e^{-s}}{(n-1)!} \quad \text{for } s \geq 0.$$

To obtain the density function $f_{Z_n}(z)$, we perform a change of variables from S to Z_n using $z = e^{-s}$ or equivalently $s = -\log(z)$. The Jacobian of this transformation is:

$$\left|\frac{ds}{dz}\right| = \frac{1}{z}.$$

Substituting, we get:

$$f_{Z_n}(z) = f_S(-\log(z)) \cdot \left|\frac{ds}{dz}\right| = \frac{(-\log(z))^{n-1}e^{\log(z)}}{(n-1)!} \cdot \frac{1}{z}.$$

Simplifying, we obtain:

$$f_{Z_n}(z) = \frac{(-\log(z))^{n-1}}{(n-1)!} \quad \text{for } 0 < z < 1.$$

Thus, we have shown that the density function for Z_n is:

$$f_n(z) = \frac{1}{(n-1)!}(-\log(z))^{n-1}.$$

Exercise 5.14. *Suppose X and Y are independent exponential random variables with the same parameter λ. Derive the probability density function (pdf) of the random variable*

$$Z = \frac{\min(X,Y)}{\max(X,Y)}.$$

Solution 5.14. *Since X and Y are independent exponential random variables with parameter λ, their pdfs are given by:*

$$f_X(x) = \lambda e^{-\lambda x} \quad \text{and} \quad f_Y(y) = \lambda e^{-\lambda y}, \quad x, y \geq 0.$$

Define $Z = \frac{\min(X,Y)}{\max(X,Y)}$. To find the pdf of Z, we analyze the probability $P(Z \leq z)$ for $0 \leq z \leq 1$.

Without loss of generality, suppose $X \leq Y$ (the case $Y \leq X$ will be symmetric). Then:

$$Z = \frac{X}{Y}.$$

Now, we need to calculate the probability $P\left(\frac{X}{Y} \leq z\right)$ under this assumption.

$$P(Z \le z) = P\left(\frac{X}{Y} \le z\right) = P(X \le zY).$$

Using the independence of X and Y, we can express this probability as:

$$P(X \le zY) = \int_0^\infty P(X \le zy) f_Y(y)\,dy.$$

Since $X \sim Exp(\lambda)$, we have $P(X \le zy) = 1 - e^{-\lambda zy}$. Thus,

$$P(Z \le z) = \int_0^\infty \left(1 - e^{-\lambda zy}\right) \lambda e^{-\lambda y}\,dy.$$

Splitting the integral, we get:

$$P(Z \le z) = \int_0^\infty \lambda e^{-\lambda y}\,dy - \int_0^\infty \lambda e^{-\lambda(1+z)y}\,dy.$$

Evaluating each term:

$$\int_0^\infty \lambda e^{-\lambda y}\,dy = 1,$$

and

$$\int_0^\infty \lambda e^{-\lambda(1+z)y}\,dy = \frac{\lambda}{\lambda(1+z)} = \frac{1}{1+z}.$$

Therefore,

$$P(Z \le z) = 1 - \frac{1}{1+z} = \frac{z}{1+z}.$$

To obtain the pdf $f_Z(z)$, differentiate $P(Z \le z)$ with respect to z:

$$f_Z(z) = \frac{d}{dz}\left(\frac{z}{1+z}\right) = \frac{1}{(1+z)^2}.$$

Thus, the pdf of $Z = \frac{\min(X,Y)}{\max(X,Y)}$ is

$$f_Z(z) = \frac{1}{(1+z)^2}, \quad 0 \le z \le 1.$$

Exercise 5.15. *A random variable Y has the pdf $f_Y(y) = Ky^{-(b+1)}$, $y \ge 2$ (and zero otherwise), where $b > 0$. This random variable is obtained as the monotonically increasing transformation $Y = g(X)$ of the random variable X with pdf e^{-x}, $x \ge 0$.*

(a) Determine K in terms of b.
(b) Determine the transformation $g(\cdot)$ in terms of b.

Solution 5.15. *We know that the total area under the probability density function $f_Y(y)$ must equal 1. Therefore, we can write:*

$$\int_2^\infty f_Y(y)\,dy = 1.$$

Substituting the expression for $f_Y(y)$:

$$\int_2^\infty Ky^{-(b+1)}\,dy = 1.$$

$$K\int_2^\infty y^{-(b+1)}\,dy = K\left[\frac{y^{-b}}{-b}\right]_2^\infty = K\left(0 + \frac{2^{-b}}{b}\right) = \frac{K}{b \cdot 2^b}.$$

Setting this equal to 1 gives us:

$$\frac{K}{b \cdot 2^b} = 1.$$

$$K = b \cdot 2^b.$$

Next, we will find the transformation $g(\cdot)$. We know that $Y = g(X)$ is a monotonically increasing transformation of X, which has the pdf:

$$f_X(x) = e^{-x}, \quad x \geq 0.$$

The cumulative distribution function (CDF) of X is:

$$F_X(x) = \int_0^x e^{-t}\, dt = 1 - e^{-x}.$$

To relate Y and X, we use the fact that $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y))$. Therefore, we can express this in terms of the CDF of X:

$$F_Y(y) = F_X(g^{-1}(y)) = 1 - e^{-g^{-1}(y)}.$$

Since we have $f_Y(y) = Ky^{-(b+1)}$, we can differentiate to find the CDF:

$$F_Y(y) = \int_2^y Kt^{-(b+1)}\, dt = K\left[\frac{t^{-b}}{-b}\right]_2^y = K\left(\frac{2^{-b}}{b} - \frac{y^{-b}}{b}\right).$$

Substituting K from part (a):

$$F_Y(y) = \frac{b \cdot 2^b}{b}\left(2^{-b} - y^{-b}\right) = 2^b\left(2^{-b} - y^{-b}\right) = 1 - \frac{2^b}{y^b}.$$

Setting this equal to $1 - e^{-g^{-1}(y)}$:

$$1 - e^{-g^{-1}(y)} = 1 - \frac{2^b}{y^b}.$$

Thus, we have:

$$e^{-g^{-1}(y)} = \frac{2^b}{y^b}.$$

Taking the natural logarithm of both sides:

$$-g^{-1}(y) = \ln\left(\frac{2^b}{y^b}\right).$$

Hence,

$$g^{-1}(y) = -b\ln\left(\frac{2}{y}\right).$$

Finally, inverting the function gives us the transformation:

$$g(x) = 2e^{-\frac{x}{b}}.$$

Exercise 5.16. *Two particles start from the same point on a two-dimensional plane and move with speed V each, such that the angle between them is uniformly distributed in $[0, 2\pi]$. Find the distribution of the magnitude of the relative velocity between the two particles.*

Solution 5.16. *Let the velocities of the two particles be represented as vectors. Let $\vec{v_1}$ and $\vec{v_2}$ be the velocity vectors of the first and second particle, respectively. We can express these velocity vectors in terms of their magnitudes and the angle θ between them:*

$$\vec{v_1} = V \begin{pmatrix} \cos(\phi_1) \\ \sin(\phi_1) \end{pmatrix}, \quad \vec{v_2} = V \begin{pmatrix} \cos(\phi_2) \\ \sin(\phi_2) \end{pmatrix}$$

where ϕ_1 and ϕ_2 are the angles of the velocities of the two particles. The angle between the two velocities is given by:

$$\theta = \phi_2 - \phi_1$$

The relative velocity $\vec{v_{rel}}$ of particle 2 with respect to particle 1 is given by:

$$\vec{v_{rel}} = \vec{v_2} - \vec{v_1} = V \begin{pmatrix} \cos(\phi_2) - \cos(\phi_1) \\ \sin(\phi_2) - \sin(\phi_1) \end{pmatrix}$$

The magnitude of the relative velocity $|\vec{v_{rel}}|$ is:

$$|\vec{v_{rel}}| = V\sqrt{(\cos(\phi_2) - \cos(\phi_1))^2 + (\sin(\phi_2) - \sin(\phi_1))^2}$$

Using the trigonometric identity for the cosine of the difference of two angles, we can rewrite the expression as follows:

$$|\vec{v_{rel}}| = V\sqrt{2 - 2\cos(\theta)} = V\sqrt{2(1 - \cos(\theta))} = V\sqrt{2}\sqrt{1 - \cos(\theta)} = V\sqrt{2}\sin\left(\frac{\theta}{2}\right)$$

Since θ is uniformly distributed in $[0, 2\pi]$, the probability density function (pdf) of θ is:

$$f_\Theta(\theta) = \frac{1}{2\pi}, \quad \textit{for } 0 \le \theta < 2\pi$$

To find the distribution of the magnitude of the relative velocity $R = |\vec{v_{rel}}|$, we can use the transformation method. We know:

$$R = V\sqrt{2}\sin\left(\frac{\theta}{2}\right)$$

To find the cumulative distribution function (CDF) of R, we can find $P(R \le r)$:

$$P(R \le r) = P\left(V\sqrt{2}\sin\left(\frac{\theta}{2}\right) \le r\right) = P\left(\sin\left(\frac{\theta}{2}\right) \le \frac{r}{V\sqrt{2}}\right)$$

Let $x = \frac{r}{V\sqrt{2}}$. The sine function maps the interval $[0, 2\pi]$ to the interval $[0, 1]$, and hence:

$$P\left(\sin\left(\frac{\theta}{2}\right) \le x\right) = P\left(\frac{\theta}{2} \le \arcsin(x)\right) = P\left(\theta \le 2\arcsin(x)\right)$$

The corresponding angles must be within the bounds of θ:

$$P\left(\theta \le 2\arcsin(x)\right) = \frac{2\arcsin(x)}{2\pi} = \frac{\arcsin(x)}{\pi}$$

Thus, the cumulative distribution function of R is:

$$F_R(r) = P(R \le r) = \frac{1}{\pi}\arcsin\left(\frac{r}{V\sqrt{2}}\right)$$

Differentiating $F_R(r)$ gives us the probability density function $f_R(r)$:

$$f_R(r) = \frac{d}{dr}F_R(r) = \frac{1}{\pi} \cdot \frac{1}{\sqrt{1 - \left(\frac{r}{V\sqrt{2}}\right)^2}} \cdot \frac{1}{V\sqrt{2}}$$

This simplifies to:

$$f_R(r) = \frac{1}{\pi V\sqrt{2}} \cdot \frac{1}{\sqrt{1 - \left(\frac{r}{V\sqrt{2}}\right)^2}}, \quad \text{for } 0 < r < V\sqrt{2}$$

Thus, the distribution of the magnitude of the relative velocity between the two particles is given by:

$$f_R(r) = \frac{1}{\pi V\sqrt{2}} \cdot \frac{1}{\sqrt{1 - \left(\frac{r}{V\sqrt{2}}\right)^2}}, \quad 0 < r < V\sqrt{2}$$

Exercise 5.17. *A point is picked uniformly from inside a unit circle. What is the density of R, the distance of the point from the center?*

Solution 5.17. *To find the density of R, the distance from the center of the unit circle, we start by noting that the area of a circle is given by the formula $A = \pi r^2$, where r is the radius of the circle. In our case, the radius is 1, so the area of the unit circle is π.*

When we choose a point uniformly inside the circle, the probability density function (pdf) must be proportional to the area element in polar coordinates. In polar coordinates, a point in the circle can be described by coordinates (r, θ), where r is the distance from the center (the value of R) and θ is the angle.

The area element in polar coordinates is given by:

$$dA = r\,dr\,d\theta$$

To find the density function of R, we need to consider how the area is distributed with respect to R. The cumulative distribution function (CDF) for R can be expressed as the probability that the distance R is less than or equal to some value r:

$$P(R \leq r) = \text{Area of the circle with radius } r = \pi r^2$$

Since R is uniformly distributed over the unit circle, the total area of the unit circle is π. Therefore, the CDF can be normalized:

$$P(R \leq r) = \frac{\pi r^2}{\pi} = r^2 \quad \text{for } 0 \leq r \leq 1$$

To find the probability density function (pdf), we take the derivative of the CDF with respect to r:

$$f_R(r) = \frac{d}{dr}P(R \leq r) = \frac{d}{dr}(r^2) = 2r \quad \text{for } 0 \leq r \leq 1$$

Thus, the density of R is given by:

$$f_R(r) = 2r \quad \text{for } 0 \leq r \leq 1$$

Exercise 5.18. *Let X and Y be independent exponentially distributed random variables with parameter 1. Find the joint density of $U = X + Y$ and $V = \frac{X}{X+Y}$, and show that V is uniformly distributed.*

Solution 5.18. *To solve this problem, we will first find the joint distribution of U and V and then show that V is uniformly distributed.*

The random variables X and Y are independent and exponentially distributed with parameter 1, so their probability density functions (pdf) are given by:

$$f_X(x) = e^{-x} \quad \text{for } x \geq 0,$$

$$f_Y(y) = e^{-y} \quad \text{for } y \geq 0.$$

Since X and Y are independent, the joint pdf is:

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = e^{-x}e^{-y} = e^{-(x+y)} \quad \text{for } x \geq 0, y \geq 0.$$

Now we define the transformation:

$$U = X + Y,$$
$$V = \frac{X}{X+Y} = \frac{X}{U}.$$

The inverse transformation is:

$$X = UV, \quad Y = U(1-V).$$

Next, we calculate the Jacobian of the transformation from (X,Y) to (U,V):

$$\begin{bmatrix} \frac{\partial X}{\partial U} & \frac{\partial X}{\partial V} \\ \frac{\partial Y}{\partial U} & \frac{\partial Y}{\partial V} \end{bmatrix} = \begin{bmatrix} V & U \\ 1-V & -U \end{bmatrix}.$$

The determinant of this Jacobian matrix is:

$$J = \begin{vmatrix} V & U \\ 1-V & -U \end{vmatrix} = -UV + U(1-V) = U.$$

Taking the absolute value gives us $|J| = U$.

Using the change of variables formula, we have:

$$f_{U,V}(u,v) = f_{X,Y}(x,y) \cdot |J| = f_{X,Y}(uv, u(1-v)) \cdot |J|.$$

Substituting for $f_{X,Y}(x,y)$:

$$f_{U,V}(u,v) = e^{-(uv+u(1-v))} \cdot u = e^{-u} \cdot u = ue^{-u} \quad \text{for } u \geq 0,\ 0 \leq v \leq 1.$$

To find the marginal distribution of V, we integrate out U:

$$f_V(v) = \int_0^\infty f_{U,V}(u,v)\,du = \int_0^\infty ue^{-u}\,du.$$

This integral is recognized as the Laplace transform of u with $s = 1$:

$$\int_0^\infty ue^{-u}\,du = 1 \quad \text{(by integration by parts or gamma function).}$$

The limits on v are $0 \leq v \leq 1$, and since we integrate over all possible values of U and find that $f_V(v)$ integrates to 1 over this range, it follows that V has a uniform distribution.

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.