Chapter 1: Probability Measure

Lecturer: Kenneth Ng

Preview

This chapter introduces fundamental concepts of abstract probability theory, which serves as a focused digest of essential topics from real analysis required in the subsequent chapters. We begin by defining a σ-algebra on a sample space Ω, which formalizes the notion of measurable events. Next, we present the definition of probability from a measure-theoretic perspective. Finally, we introduce random variables as measurable functions defined on the probability space.

Key topics in this chapter:

1. σ-algebra;
2. Probability as a measure;
3. Random variables as measurable functions;
4. Distributions of random variables.

1 Sigma Algebra

Let Ω be a *sample space*, and let $\mathcal{F}$ be a collection of subsets of Ω, which represents the observable information in the model, consisting of events $A \subseteq \Omega$ with accessible probabilities. This collection $\mathcal{F}$ must satisfy certain desirable properties. For example, if we can assign a probability to an event A, we should also be able to assign a probability to $A^c := \Omega \setminus A$. This motivates the requirement that $A \in \mathcal{F}$ implies $A^c \in \mathcal{F}$. More generally, $\mathcal{F}$ must be rich enough to support operations needed for a coherent probability assignment.

Definition 1.1 A collection $\mathcal{F}$ is said to be a ***σ-algebra*** (a.k.a. ***σ-field***) if it satisfies the following properties:

1. $\Omega \in \mathcal{F}$;
2. (closed under complementation) if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$;
3. (closed under countable union) for any $A_1, A_2, \cdots \in \mathcal{F}$, $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

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

The first property ensures that we know the probability of the entire sample space, which is naturally 1. The second property has been discussed above, reflecting the need for complements of events to also be measurable. The third property states that if we know the probabilities of a sequence of events, we should also be able to determine the probability that the sequence of events occurring simultaneously.

Some properties of a σ-algebra:

1. $\varnothing \in \mathcal{F}$.
2. By De Morgan's law, Properties 2 and 3 imply that a σ-algebra is closed under countable union: for any $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$, $\cap_{n=1}^{\infty} A_n \in \mathcal{F}$.

Example 1.1

1. The collection $\mathcal{F} = \{\varnothing, \Omega\}$ is called a *trivial σ-algebra*.
2. For any $A \subseteq \Omega$, $\mathcal{F} = \sigma(A) := \{\varnothing, A, A^c, \Omega\}$ is the σ-algebra generated by the set A.
3. In general, for any collection of subsets $\mathcal{C} \subseteq 2^{\Omega}$, where 2^{Ω} is the *power set* of Ω which contains all subsets of Ω, we denote by $\sigma(\mathcal{C})$ the *smallest σ-algebra generated by $\mathcal{C}$*.

Example 1.2 A coin is tossed twice and the outcome is recorded as H for a head, and T for a tail. The sample space is then $\Omega = \{HH, HT, TH, TT\}$. Let $\mathcal{C} = \{\{HH\}, \{HT, TH\}, \{TT\}\}$. Write down $\mathcal{F} = \sigma(\mathcal{C})$.

Solution. Since $\mathcal{C}$ consists of 3 pairwise disjoint subsets of Ω, $\mathcal{F}$ will contain $2^3 = 8$ subsets including Ω and $\varnothing$, which is given by

$$\mathcal{F} = \{\varnothing, \Omega, \{HH\}, \{HT, TH\}, \{TT\}, \{HH, HT, TH\}, \{HH, TT\}, \{HT, TH, TT\}\}.$$

□

Example 1.3 Let $d \in \mathbb{N}$ and let $\Omega = \mathbb{R}^d$. The **Borel σ-algebra**, denoted by $\mathcal{B}(\mathbb{R}^d)$, is the smallest σ-algebra containing all open subsets of $\mathbb{R}^d$, i.e., $\mathcal{B}(\mathbb{R}^d) = \sigma(\mathcal{O})$, where $\mathcal{O}$ is the collection of all open subsets in $\mathbb{R}^d$. It is the collection of subsets of $\mathbb{R}^d$ generated by open sets under countable unions, countable intersections, and complementation. Any set in $\mathcal{B}(\mathbb{R}^d)$ is called a ***Borel set***. In particular, any closed or open set in $\mathbb{R}^d$ is a Borel set.

The next result shows that the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ can be generated by countably many one-sided closed intervals of the form $[-\infty, a)$. This allows us to reduce many constructions and proofs to countable operations, where the behavior of measure-zero or measure-one sets is well controlled through countable additivity.

Theorem 1.1 The Borel σ-algebra of $\mathbb{R}$ is generated by the collection $\mathcal{C}$ (i.e., $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C})$), where

$$\mathcal{C} = \{(-\infty, a] : a \in \mathbb{Q}\}.$$

Proof. Let $\mathcal{I}$ be the collection of all open intervals. Using the density of $\mathbb{Q}$ in $\mathbb{R}$, one can show that every open set is a countable union of open intervals. Hence, $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I})$.

Next, we show that $\mathcal{I} \subseteq \sigma(\mathcal{C})$. For any $a, b \in \mathbb{R}$, $b > a$, we can find sequences of rational numbers $(a_n)_n$ and $(b_n)_n$ such that $b_n > a_n$, $a_n \downarrow a$ and $b_n \uparrow b$. Using this, we can express

$$(a, b) = \bigcup_{n=1}^{\infty} (a_n, b_n] = \bigcup_{n=1}^{\infty} \left((-\infty, b_n] \cap (-\infty, a_n]^c\right),$$

which verifies the claim. Hence, $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I}) \subseteq \sigma(\mathcal{C})$. On the other hand, each set in $\mathcal{C}$ is a closed set, and thus a Borel set. This implies $\mathcal{C} \subseteq \mathcal{B}(\mathbb{R})$, and thus $\sigma(\mathcal{C}) \subseteq \mathcal{B}(\mathbb{R})$. Therefore, $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$.

□

Proposition 1.2 Let $\mathcal{F}$ and $\mathcal{G}$ be two σ-algebras. Then, $\mathcal{F} \cap \mathcal{G}$ is again a σ-algebra. However, $\mathcal{F} \cup \mathcal{G}$ is not necessarily a σ-algebra.

Proof. We verify that $\mathcal{F} \cap \mathcal{G}$ satisfies the 3 properties of a σ-algebra if $\mathcal{F}$ and $\mathcal{G}$ are both σ-algebras:

1. Since $\mathcal{F}$ and $\mathcal{G}$ are σ-algebras, $\Omega \in \mathcal{F}$ and $\Omega \in \mathcal{G}$, which implies $\Omega \in \mathcal{F} \cap \mathcal{G}$.
2. For any $A \in \mathcal{F} \cap \mathcal{G}$, so that $A \in \mathcal{F}$ and $A \in \mathcal{G}$, we have $A^c \in \mathcal{F}$ and $A^c \in \mathcal{G}$ (the second property of σ-algebras), and so $A^c \in \mathcal{F} \cap \mathcal{G}$.
3. For any collection of subsets $\{A_n\}_{n=1}^{\infty}$ such that $A_n \in \mathcal{F} \cap \mathcal{G}$ for all $n > 0$, we have $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$ and $\cup_{n=1}^{\infty} A_n \in \mathcal{G}$ (the third property of σ-algebras), whence $\cup_{n=1}^{\infty} A_n \in \mathcal{F} \cap \mathcal{G}$.

To show that $\mathcal{F} \cup \mathcal{G}$ might not be a σ-algebra, we give the following counter-example: define $\Omega := \{a, b, c\}$, $\mathcal{F} := \{\varnothing, \{a\}, \{b, c\}, \Omega\}$, and $\mathcal{G} := \{\varnothing, \{b\}, \{a, c\}, \Omega\}$. It is easy to check that $\mathcal{F}$ and $\mathcal{G}$ are σ-algebras (exercise). However,

$$\mathcal{F} \cup \mathcal{G} = \{\varnothing, \{a\}, \{b\}, \{a, c\}, \{b, c\}, \Omega\}$$

is not a σ-algebra (exercise). □

If $\mathcal{F}$ and $\mathcal{G}$ are two σ-algebras on the same sample space Ω, we denote by $\mathcal{F} \vee \mathcal{G}$ the smallest σ-algebra that contains both $\mathcal{F}$ and $\mathcal{G}$, i.e., $\mathcal{F} \vee \mathcal{G} = \sigma(\mathcal{F} \cup \mathcal{G})$.

2 Probability Measure

A measure defined on a measurable space $(\Omega, \mathcal{F})$ is a way to assign a numerical value to an event $A \in \mathcal{F}$ that essentially quantifies their "size" or "extent". In particular, a probability measure is a special class of measure, which satisfies the following properties:

Definition 2.1 A ***probability measure*** $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$ is a mapping $A \in \mathcal{F} \mapsto \mathbb{P}(A)$ that satisfies the following properties:

1. for any $A \in \mathcal{F}$, $\mathbb{P}(A) \in [0, 1]$;
2. $\mathbb{P}(\Omega) = 1$;
3. (countable additivity) let $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ be a countable collection of pairwise disjoint sets, i.e., $A_i \cap A_j = \varnothing$ for $i \neq j$, we have

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

The tuple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a ***probability space***.

Remark 2.1.

1. If $\mu : \mathcal{F} \to [0, \infty)$ satisfies only the third property and $\mu(\varnothing) = 0$, then μ is called a ***measure***.
2. If μ is a measure, then we call the tuple $(\Omega, \mathcal{F}, \mu)$ a measure space.
3. If $\mu(\Omega) < \infty$, then μ is called a ***finite measure***.
4. If there exists a collection $A_1, A_2, \dots \in \mathcal{F}$ with $\cup_{n=1}^{\infty} A_n = \Omega$, and $\mu(A_n) < \infty$ for all n, then μ is said to be a ***σ-finite measure***.
5. Probability measures $\subset$ finite measures $\subset$ σ-finite measures.

The following are some fundamental properties of a probability measure.

Theorem 2.2 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $A, B \in \mathcal{F}$, and $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$. The following properties hold:

1. (monotonicity). If $A \subseteq B$, $\mathbb{P}(A) \leq \mathbb{P}(B)$.
2. (sub-additivity). If $A \subseteq \cup_{n=1}^{\infty} A_n$, $\mathbb{P}(A) \leq \mathbb{P}(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.
3. (continuity from below). If $A_n \uparrow A$ (i.e., $A_1 \subseteq A_2 \subseteq \cdots$ and $\cup_{n=1}^{\infty} A_n = A$). Then $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$.
4. (continuity from above). If $A_n \downarrow A$ (i.e., $A_1 \supset A_2 \supset \cdots$ and $\cap_{n=1}^{\infty} A_n = A$). Then $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$.

Proof. 1. Let $C := B \backslash A = B \cap A^c$. Then $A \cap C = \varnothing$ and $A \cup C = B$. By the countable additivity and non-negativity of probability measures,

$$\mathbb{P}(A) \leq \mathbb{P}(A) + \mathbb{P}(C) = \mathbb{P}(B).$$

2. By the first property we have $\mathbb{P}(A) \leq \mathbb{P}(\cup_{n=1}^{\infty} A_n)$. Hence, it suffices to show that $\mathbb{P}(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$. To this end, let $B_1 := A_1$, and for $n = 2\ldots$, let $B_n := A_n \backslash \cup_{k=1}^{n-1} A_k$. Then, $B_n \cap B_m = \varnothing$ for $n \neq m$, and $\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n$. By the countable additivity of probability measures and noting that $B_n \subseteq A_n$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

3. Let $B_1 := A_1$, and for $n \geq 1$, let $B_n := A_n \backslash A_{n-1}$. Then, $\{B_n\}_{n=1}^{\infty}$ is pairwise disjoint, $\cup_{k=1}^{n} B_k = A_n$, and $\cup_{k=1}^{\infty} B_k = A$. Hence,

$$\mathbb{P}(A_n) = \sum_{k=1}^{n} \mathbb{P}\left(B_k\right) \uparrow \sum_{k=1}^{\infty} \mathbb{P}\left(B_k\right) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} B_k\right) = \mathbb{P}(A).$$

4. For any $n \geq 1$, let $B_n := A_n^c$. Then, $B_1 \subseteq B_2 \subseteq \cdots$ and $B_n \uparrow B := A^c$. Using Property 3, we have $1 - \mathbb{P}(A_n) = \mathbb{P}(B_n) \uparrow \mathbb{P}(B) = 1 - \mathbb{P}(A)$, and hence $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$.

$\square$

Example 2.1 (Lebesgue measure) Let $\Omega = [0,1]$ and $\mathcal{F} = \mathcal{B}([0,1])$ be the Borel σ-algebra. Define the ***Lebesgue measure***[a] λ on $(\Omega, \mathcal{F})$ by, for any $0 \leq a \leq b \leq 1$,

$$\lambda([a,b]) = b - a.$$

Compute $\lambda(\{a\})$ and $\lambda((a,b])$.

Solution. The singleton $\{a\}$ can be written as $\{a\} = [a,a]$. Hence, $\lambda(\{a\}) = \lambda([a,a]) = a - a = 0$. In other words, a point in $[0,1]$ has no "length". Using this, we can write $(a,b] = [a,b] \cap \{a\}^c$, whence $\lambda((a,b]) = \lambda([a,b]) - \mathbb{P}(\{a\}) = b - a$. $\square$

[a] The proof of Lebesgue measure defined on $[0,1]$ is indeed a probability measure is out of the scope of the course. More generally, we can define the Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ analogously by $\lambda([a,b]) = b - a$ for any $a, b \in \mathbb{R}$, $a \leq b$.

An event $A \in \mathcal{F}$ is said to occur ***almost surely*** (a.s.) if $\mathbb{P}(A) = 1$. A set $N \subseteq \mathcal{F}$ is said to be a $\mathbb{P}$-*null set* if $\mathbb{P}(N) = 0$. In general, if μ is a measure and $\mu(A^c) = 0$, then we say that A occurs ***almost everywhere*** (μ-a.e.). To meaningfully talk about events that happen with probability zero or one, we complete the σ-algebra $\mathcal{F}$ so that all such events are guaranteed to be measurable. The resulting σ-algebra is called the $\mathbb{P}$-completion:

Definition 2.2 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then, the $\mathbb{P}$-completion of $\mathcal{F}$ is $\mathcal{F}' := \mathcal{F} \vee \mathcal{N}$, where

$$\mathcal{N} := \{A \subseteq \Omega : A \subseteq N,\ N \in \mathcal{F}, \mathbb{P}(N) = 0\}$$

is the subsets of all $\mathbb{P}$-null sets in $\mathcal{F}$.

3 Random Variables

A random variable X is a measurable mapping from the sample space Ω to a suitable target space (often $\mathbb{R}^d$) that assigns a value $X(\omega)$ to each outcome $\omega \in \Omega$, reflecting the statistical quantity of interest. In order to describe the probability of $\{X \in A\}$, the pre-image of this event must lie in the information set (i.e., $\mathcal{F}$) of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. This motivates the definition of a random variable:

Definition 3.1 Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be measurable spaces. An ***$\mathcal{F}$-measurable function*** is a map $X : \Omega \to S$ such that, for any $A \in \mathcal{S}$,

$$X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}.$$

A ***random variable*** is a measurable function defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

When the σ-algebra $\mathcal{S}$ of the target space S is generated by a collection of subsets $\mathcal{A}$ of S. To show that $X : \Omega \to S$ is a random variable, it suffices to check that $X^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{A}$. In particular:

- If $S = \{x_1, \ldots, x_n\}$ and $\mathcal{S} = 2^S$, it suffices to check that $X^{-1}(\{x_i\}) \in \mathcal{F}$ for all $i = 1, \ldots, n$. A measurable function (resp. random variable) that takes finitely many values is also called a *simple function* (resp. *simple random variable*).
- If $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, it suffices to check $X^{-1}([-\infty, a)) \in \mathcal{F}$ for all $a \in \mathbb{Q}$.

Proposition 3.1 Let S be a set, $\mathcal{A}$ be a collection of subsets of S, and $\mathcal{S} = \sigma(\mathcal{A})$ be the σ-algebra generated by $\mathcal{A}$. Then, $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ is measurable if and only if $X^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{A}$.

Proof. Let $A \in \mathcal{A}$. Since $\mathcal{S} = \sigma(\mathcal{A})$, we have $A \in \mathcal{S}$. By the measurability of X, we have $X^{-1}(A) \in \mathcal{F}$.

Now suppose that $X^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{A}$. Define

$$\mathcal{G} := \{A \subseteq S :\ X^{-1}(A) \in \mathcal{F}\}.$$

By construction and the statement's assumption, we have $\mathcal{A}] \subseteq \mathcal{G}$.

We claim $\mathcal{G}$ is a σ-algebra on S:

1. $S \in \mathcal{G}$ since $X^{-1}(S) = \Omega \in \mathcal{F}$.
2. If $A \in \mathcal{G}$, we have $X^{-1}(A) \in \mathcal{G}$. Then, $X^{-1}(A^c) = \Omega \setminus X^{-1}(A) \in \mathcal{F}$ since $\mathcal{F}$ is a σ-algebra. This implies $A^c \in \mathcal{G}$.
3. If $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{G}$, then $X^{-1}(\bigcup_n A_n) = \bigcup_n X^{-1}(A_n) \in \mathcal{F}$, so $\bigcup_n A_n \in \mathcal{G}$.

Since $\mathcal{G}$ is a σ-algebra containing $\mathcal{A}$, we have $\mathcal{S} = \sigma(\mathcal{A}) \subseteq \mathcal{G}$. Therefore, for every $B \in \mathcal{S}$, we have $B \in \mathcal{G}$ and hence $X^{-1}(B) \in \mathcal{F}$, i.e., X is $\mathcal{F}/\mathcal{S}$-measurable.

□

Example 3.1 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A \in \mathcal{F}$. The ***indicator function*** $\mathbb{1}_A : \Omega \to \{0, 1\}$ is a random variable given by

$$\mathbb{1}_A(x) = \begin{cases} 1, & \text{if } x \in A; \\ 0, & \text{if } x \notin A. \end{cases}$$

Example 3.2 Continuing from Example 1.2, determine whether the following are measurable functions on $(\Omega, \mathcal{F})$ to $(\mathbb{N}_0, 2^{\mathbb{N}_0})$:

1. X returns the number of heads in the two tosses;
2. Y returns 1 if the first toss is a head, and 0 otherwise.

Solution.

1. X can take the values 0, 1, or 2.

$$X^{-1}(\{0\}) = \{TT\}, \; X^{-1}(\{1\}) = \{HT, TH\}, \; X^{-1}(\{2\}) = \{HH\}.$$

   $\mathcal{F}$ contains all these subsets, whence X is a measurable function.
2. Y can take the values 0 or 1. Note that $Y^{-1}(\{0\}) = \{TT, TH\} \notin \mathcal{F}$. Hence, Y is not measurable.

□

The mapping Y in Example 3.2 is not a random variable with respect to $\mathcal{F}$, as $\mathcal{F}$ is not sufficiently fine to capture the information about the outcome of the first coin toss. To make Y measurable, a finer σ-algebra is required. The smallest such σ-algebra is defined below:

Definition 3.2 Let $X : \Omega \to S$. The ***σ-algebra generated by*** X, denoted by $\sigma(X)$, is defined as

$$\sigma(X) := \{X^{-1}(A) : A \in \mathcal{S}\}.$$

As shown in the next result, $\sigma(X)$ is indeed a σ-algebra. By constructions, $\sigma(X)$ is the smallest σ-algebra such that X is a measurable function.

Theorem 3.2 $\sigma(X)$ is a σ-algebra. In addition, for any σ-algebra $\mathcal{F}$ of Ω such that X is $\mathcal{F}$-measurable, we have $\sigma(X) \subseteq \mathcal{F}$.

Proof. We first show that $\sigma(X)$ is a σ-algebra by verifying the three properties:

1. $\Omega \in \sigma(X)$: Take $A = S$, $\Omega = X^{-1}(S) \in \sigma(X)$.
2. $B \in \sigma(X) \Rightarrow B^c \in \sigma(X)$: Let $A \in \mathcal{S}$ such that $B = X^{-1}(A) \in \sigma(X)$. Then, $B^c = X^{-1}(A^c) \in \sigma(X)$ since $A^c \in \mathcal{S}$.
3. $B_1, B_2, \dots \subseteq \sigma(X) \Rightarrow \cup_{n=1}^{\infty} B_n \in \sigma(X)$: Let $A_n \in \mathcal{S}$ such that $B_n = X^{-1}(A_n) \in \sigma(X)$. Since $\mathcal{S}$ is a σ-algebra, $\cup_{n=1}^{\infty} A_n \in \mathcal{S}$, and thus

$$\sigma(X) \ni X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(A_n) = \bigcup_{n=1}^{\infty} B_n.$$

Therefore, $\sigma(X)$ is a σ-algebra.

Now, suppose that X is $\mathcal{F}$-measurable. For any $B \in \sigma(X)$, we have $B = X^{-1}(A) \in \mathcal{F}$ since X is $\mathcal{F}$-measurable. Hence, $\sigma(X) \subseteq \mathcal{F}$. $\square$

Example 3.3 Continuing from Example 3.2, find $\sigma(Y)$.

Solution. Note that $Y^{-1}(\{0\}) = \{TT, TH\}$ and $Y^{-1}(\{1\}) = \{HT, HH\}$. Hence, $\sigma(Y) = \{\varnothing, \{TT, TH\}, \{HT, HH\}, \Omega\}$. One can also check that $\mathcal{F} = \sigma(X)$. $\square$

The following result says that composition of measurable functions is again measurable.

Theorem 3.3 Suppose that $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ and $f : (S, \mathcal{S}) \to (U, \mathcal{U})$ are measurable functions. Then, $f \circ X : (\Omega, \mathcal{F}) \to (U, \mathcal{U})$ is also measurable.

Proof. It suffices to show that $(f \circ X)^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{U}$. Note that by the measurability of f, $f^{-1}(A) \in \mathcal{S}$ for any $A \in \mathcal{U}$. Using this and the measurability of X, we also have $X^{-1}(f^{-1}(A)) \in \mathcal{F}$ for any $A \in \mathcal{S}$. The proof is thus complete by noting that $(f \circ X)^{-1}(A) = X^{-1}(f^{-1}(A))$. $\square$

We also have the following generalizations, and the proof is omitted herein.

Theorem 3.4 Suppose that $X_1, \dots, X_n : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}(\mathcal{R}^n))$, and $f : (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are measurable functions. Then, $f(X_1, \dots, X_n) : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathbb{R}^n)$ is also measurable.

Theorem 3.4 implies that if $\{X_i\}_{i=1}^n$ is a sequence of random variables, then $\sum_{i=1}^n X_i$, $\prod_{i=1}^n X_i$ are also random variables. In addition, as shown below, the supremum and infinium of a sequence of random variables are again random variables. The proof is relegated to the appendix for interested readers.

Theorem 3.5 If $X_1, \dots, X_n$ are $\mathbb{R}$-valued random variables. Then, the following are also random variables:

$$\inf_n X_n, \ \sup_n X_n, \ \liminf_n X_n, \ \limsup_n X_n.$$

Here,

$$\liminf_n X_n := \lim_{n\to\infty} \left(\inf_{m \geq n} X_m \right) = \sup_n \left(\inf_{m \geq n} X_m \right),$$
$$\limsup_n X_n := \lim_{n\to\infty} \left(\sup_{m \geq n} X_m \right) = \inf_n \left(\sup_{m \geq n} X_m \right).$$

4 Distributions

The distribution P_X of a random variable X describes how probability is assigned to the possible values that X can take. In measure theory, P_X is also called the *pushforward measure* of $\mathbb{P}$ by X.

Definition 4.1 Let X be a random variable from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the target space $(S, \mathcal{S})$. The ***distribution*** (a.k.a. ***law***) P_X of X is a probability measure on $(S, \mathcal{S})$ defined by

$$P_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega : X(\omega) \in A\}), \ A \in \mathcal{S}.$$

The proof that P_X is a probability measure on the space $(S, \mathcal{S})$ is left as an exercise. If X takes finitely or countably many values, then

$$P_X(A) = \mathbb{P}\left(\bigcup_{x \in A} \{\omega : X(\omega) = x\} \right) = \sum_{x \in A} \mathbb{P}(X^{-1}(\{\boldsymbol{x}\})).$$

4.1 Distribution Function

If X takes values in $\mathbb{R}$, its distribution is characterized by the probability over the Borel sets, which are in particular generated by the intervals of the form $[-\infty, a)$, $a \in \mathbb{R}$. This motivates us to define the distribution function of a $\mathbb{R}$-valued random variable.

Definition 4.2 Let X be a $\mathbb{R}$-valued random variable. The ***distribution function*** of X is defined as

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

The following lists the properties of a distribution function.

Theorem 4.1 The distribution function F_X of a $\mathbb{R}$-valued random variable X must satisfy the following properties:

1. F_X is non-decreasing;
2. $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$;
3. F_X is right-continuous, i.e., $\lim_{x\downarrow x_0} F_X(x) = F_X(x_0)$;
4. $F_X(x_0^-) := \lim_{x\uparrow x_0} F_X(x) = \mathbb{P}(X < x_0)$;
5. $\mathbb{P}(X = x) = F_X(x) - F_X(x^-) = F_X(x) - \mathbb{P}(X < x)$.

Proof.

1. For any $x \leq y$, $\{X \leq x\} \subseteq \{X \leq y\}$, and so $F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y) = F_X(y)$.
2. $\{X \leq x\} \downarrow \varnothing$ as $x \downarrow -\infty$, and $\{X \leq x\} \uparrow \Omega$ as $x \uparrow \infty$. Hence, the result follows from Statements 3 and 4 of Theorem 2.2.
3. Since $\{X \leq x\} \downarrow \{X \leq x_0\}$ as $x \downarrow x_0$, the result follows from Statement 4 of Theorem 2.2.
4. Since $\{X \leq x\} \uparrow \{X < x_0\}$ (NOT $\{X \leq x_0\}$!), the result follows from Statement 3 of Theorem 2.2.
5. Since $\{X = x\} \cup \{X < x\} = \{X \leq x\}$, and $\{X = x\} \cap \{X < x\} = \varnothing$, we have $F_X(x) = \mathbb{P}(X = x) + \mathbb{P}(X < x) = F_X(x^-) + \mathbb{P}(X < x)$.

□

The next result shows that Properties 1-3 in Theorem 4.1 characterize a distribution function. We omit the proof herein.

Theorem 4.2 Let $F : \mathbb{R} \to [0, 1]$ be a function that satisfies Properties 1-3 of Theorem 4.1. Then, F is the distribution function of some random variable X.

4.2 Density Function and the Radon-Nikodym Theorem

In basic probability courses, we learn that a continuous random variable admits a probability density function, which allows us to compute probabilities and other statistical quantities via integration. We now generalize this idea in a measure-theoretic setting. To do so, we first introduce the notion of absolute continuity, which allows us to compare the "fineness" of two measures:

Definition 4.3 Let $(S, \mathcal{S})$ be a measurable space, and let μ and ν be two measures defined on $(S, \mathcal{S})$. We say that μ is ***absolutely continuous*** with respect to ν, denoted $\mu \ll \nu$, if for every set $A \in \mathcal{S}$, $\nu(A) = 0$ implies $\mu(A) = 0$.

The following provides a general condition a which a $\mathbb{R}$-valued random variable admits a density function.

Theorem 4.3 (Radon–Nikodym) Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be a real-valued random variable, and let P_X denote the distribution of X. If P_X is absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}$ (i.e., $P_X \ll \lambda$), then there exists a measurable function $f_X : \mathbb{R} \to [0, \infty)$, called the ***probability density function (pdf)*** of X, such that

$$P_X(B) = \mathbb{P}(X \in B) = \int_B f_X(x)\, dx, \quad \text{for all Borel sets } B \in \mathcal{B}(\mathbb{R}).$$

Here are some remarks regarding Theorem 4.3:

1. The integral involving the density function is understood in the *Lebesgue* sense, which will be discussed in detail in the next chapter.

2. The Radon–Nikodym theorem can be interpreted as a *change of measure* result: it relates the probability measure $\mathbb{P}$ to the Lebesgue measure λ through a density function. We will revisit this concept in the context of risk-neutral measures and Girsanov's theorem.

3. The density function f_X is also referred to as the ***Radon–Nikodym derivative*** of $\mathbb{P}$ with respect to λ, and is denoted by

$$f_X(x) = \frac{d\mathbb{P}}{d\lambda}(x).$$

   In particular, when the distribution function F_X is differentiable, $f_X = \frac{d}{dx} F_X$.

A Proof of Theorem 3.5

We first show that $\sup_n X_n$ is a random variable. Note that for any $x \in \mathbb{R}$,

$$\left\{ \sup_n X_n \leq x \right\} = \bigcap_n \{X_n \leq x\} \in \mathcal{F},$$

since each $\{X_n \leq x\} \in \mathcal{F}$ by the measurability of X_n. Note that the equivalence of the two sets holds since $\sup_n X_n(\omega) \leq x \iff X_k(\omega) \leq x$ for all $k \in \mathbb{N}$. Since any Borel set is generated by the one-sided closed set $(-\infty, x]$, we conclude that $\sup_n X_n$ is a random variable; see Proposition 3.1. To show that $\inf_n X_n$ is a random variable, it suffices to note that

$$\inf_n X_n = -\sup_n(-X_n).$$

To show that $\liminf_n X_n$ is a random variable, let $Y_n := \inf_{m\geq n} X_m$, which is a random variable using the proven fact about the infinium of random variables. Hence, $\liminf_n X_n = \sup_n Y_n$, which is a random variable, again using the proven fact about the supremum of random variables. The fact that $\limsup_n X_n$ is a random variable can be shown in a similar manner.