Probability Theory I Notes

Instructor: Prof. Konstantinos Spiliopoulos

Student: Wancheng Lin

Contents

---

1 Unit 1: Foundations of Probability

Sets and Elements

Definition 1.1 (Set). *A* set *is a collection of objects called* elements. *We write $x \in A$ if x belongs to A, and $x \notin A$ otherwise.*

Example (Basic Examples of Sets).This example illustrates different kinds of sets:

- $A = \{1, 2, 3\}$ (finite set),
- $\mathbb{N} = \{0, 1, 2, \dots\}$ (countable infinite set),
- $\mathbb{R}$ (uncountable set),
- $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$ (the closed unit disk in $\mathbb{R}^2$).

Definition 1.2 (Subset). *For sets A, B, we say A is a* subset *of B (denoted $A \subseteq B$) if*

$$\forall x \ (x \in A \implies x \in B).$$

Basic Set Operations

For $A, B \subseteq \Omega$, the usual operations are:

$$A \cup B = \{x : x \in A \text{ or } x \in B\}, \quad A \cap B = \{x : x \in A \text{ and } x \in B\},$$

$$A^c = \Omega \setminus A, \quad A \setminus B = \{x \in A : x \notin B\}, \quad A \triangle B = (A \setminus B) \cup (B \setminus A).$$

Theorem 1.1 (De Morgan's Laws). *For $\{A_n\}_{n=1}^{\infty} \subseteq \Omega$,*

$$\left(\bigcup_{n=1}^{\infty} A_n\right)^c = \bigcap_{n=1}^{\infty} A_n^c, \qquad \left(\bigcap_{n=1}^{\infty} A_n\right)^c = \bigcup_{n=1}^{\infty} A_n^c.$$

Countable and Uncountable Sets

Definition 1.3 (Countable). *A set A is* countable *if either it is finite or there exists a bijection $f : A \to \mathbb{N}$.*

Example (Countable vs Uncountable Sets).We contrast common countable and uncountable sets:

- Countable: $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$.
- Uncountable: $(0, 1), \mathbb{R}$, and the set of infinite binary sequences (Cantor's diagonal argument).

Diagonal Argument for Uncountability

Theorem 1.2 (Cantor's Diagonal Argument). *The set of all infinite binary sequences*

$$\{(a_n)_{n=1}^{\infty} : a_n \in \{0, 1\}\}$$

is uncountable.

Proof. Assume, for contradiction, that the set of infinite 0–1 sequences can be listed as

$$s^{(1)}, s^{(2)}, s^{(3)}, \dots$$

where each $s^{(i)} = (a_1^{(i)}, a_2^{(i)}, a_3^{(i)}, \dots)$.

Now construct a new sequence $t = (b_1, b_2, b_3, \dots)$ by choosing

$$b_n = \begin{cases} 0, & a_n^{(n)} = 1, \\ 1, & a_n^{(n)} = 0. \end{cases}$$

That is, b_n is defined to be different from the n-th entry of the n-th sequence. By construction, t differs from $s^{(n)}$ in the n-th coordinate for every n. Thus t is not in the list, contradicting the assumption that all sequences were listed. Therefore, the set of infinite binary sequences is uncountable. □

Remark. The same diagonalization argument shows that the interval $(0, 1)$ is uncountable, since each binary sequence corresponds to the binary expansion of a number in $(0, 1)$.

Countability of the Rational Numbers

Theorem 1.3. *The set of rational numbers $\mathbb{Q}$ is countable.*

Proof. Every rational number can be expressed as $\frac{p}{q}$ with $p \in \mathbb{Z}$ and $q \in \mathbb{N}$. Consider the infinite array

$$\begin{array}{cccc} \frac{1}{1} & \frac{2}{1} & \frac{3}{1} & \cdots \\ \frac{1}{2} & \frac{2}{2} & \frac{3}{2} & \cdots \\ \frac{1}{3} & \frac{2}{3} & \frac{3}{3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

We can enumerate these fractions by traversing the array diagonally:

$$\frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{1}{3}, \frac{2}{2}, \frac{3}{1}, \frac{4}{1}, \frac{3}{2}, \frac{2}{3}, \frac{1}{4}, \ldots$$

This procedure produces a sequence containing every positive rational number infinitely many times. By discarding duplicates (e.g. $\frac{2}{2} = 1$, $\frac{3}{3} = 1$, etc.), we obtain a bijection between $\mathbb{N}$ and $\mathbb{Q}^+$, the set of positive rationals.

Finally, we can interleave negatives:

$$0, 1, -1, \tfrac{1}{2}, -\tfrac{1}{2}, 2, -2, \tfrac{2}{3}, -\tfrac{2}{3}, \ldots$$

to cover all of $\mathbb{Q}$.

Hence, $\mathbb{Q}$ is countable. □

Remark. This construction illustrates that while $\mathbb{Q}$ is dense in $\mathbb{R}$, it can still be arranged into a list. In contrast, $\mathbb{R}$ is uncountable by Cantor's diagonal argument.

Sequences of Sets

Definition 1.4 (Monotone Sequences of Sets)**.** *Consider that*

- ***Increasing:*** $A_1 \subseteq A_2 \subseteq \ldots$. *Then*
$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$
- ***Decreasing:*** $A_1 \supseteq A_2 \supseteq \ldots$. *Then*
$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

Limit Inferior and Limit Superior of Sets

Definition 1.5 (Limit Inferior of Sets)**.** *Given a sequence of sets $(A_n)_{n\geq 1}$, the* limit inferior *is*

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \{\, x : \exists n \text{ s.t. } \forall m \geq n,\ x \in A_m \,\}.$$

This is the set of points that eventually belong to all A_m*, i.e. belong to all but finitely many sets.*

Definition 1.6 (Limit Superior of Sets)**.** *The* limit superior *is*

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\, x : x \in A_m \text{ for infinitely many } m \,\}.$$

This is the set of points that occur infinitely often *in the sequence.*

Remark. Keep in mind that

- $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$ always holds.
- $\bigcap_{m=n}^{\infty} A_m$ is increasing in n, while $\bigcup_{m=n}^{\infty} A_m$ is decreasing in n.

Assignment

Which of the following statements are (not) true?

$$\limsup_{n\to\infty}(A_n \cup B_n) = \big(\limsup_{n\to\infty} A_n\big) \cup \big(\limsup_{n\to\infty} B_n\big), \tag{a}$$

$$\limsup_{n\to\infty}(A_n \cap B_n) = \big(\limsup_{n\to\infty} A_n\big) \cap \big(\limsup_{n\to\infty} B_n\big), \tag{b}$$

$$\liminf_{n\to\infty}(A_n \cup B_n) = \big(\liminf_{n\to\infty} A_n\big) \cup \big(\liminf_{n\to\infty} B_n\big), \tag{c}$$

$$\liminf_{n\to\infty}(A_n \cap B_n) = \big(\liminf_{n\to\infty} A_n\big) \cap \big(\liminf_{n\to\infty} B_n\big). \tag{d}$$

Moreover, if $A_n \to A$ and $B_n \to B$ (i.e. both limsup and liminf equal), then

$$A_n \cup B_n \to A \cup B \quad \text{as } n \to \infty, \qquad A_n \cap B_n \to A \cap B \quad \text{as } n \to \infty.$$

I. We first conclude that (a) and (d) in the following two are **true**:

$$\limsup_{n\to\infty}(A_n \cup B_n) \;=\; \big(\limsup_{n\to\infty} A_n\big) \cup \big(\limsup_{n\to\infty} B_n\big) \text{ and } \liminf_{n\to\infty}(A_n \cap B_n) \;=\; \big(\liminf_{n\to\infty} A_n\big) \cap \big(\liminf_{n\to\infty} B_n\big).$$

II. But the (b) and (c) in the following two are **not true**:

$$\limsup_{n\to\infty}(A_n \cap B_n) \;=\; \big(\limsup_{n\to\infty} A_n\big) \cap \big(\limsup_{n\to\infty} B_n\big) \text{ and } \liminf_{n\to\infty}(A_n \cup B_n) \;=\; \big(\liminf_{n\to\infty} A_n\big) \cup \big(\liminf_{n\to\infty} B_n\big).$$

In general, they should instead follow the relationships below, denoted as equation (b^*) and (c^*):

$$\limsup_{n\to\infty}(A_n \cap B_n) \;\subset\; \big(\limsup_{n\to\infty} A_n\big) \cap \big(\limsup_{n\to\infty} B_n\big) \text{ and } \liminf_{n\to\infty}(A_n \cup B_n) \;\supset\; \big(\liminf_{n\to\infty} A_n\big) \cup \big(\liminf_{n\to\infty} B_n\big).$$

Proof. **(a)** This equation is true, and we will prove it by establishing by **double inclusion**:

Firstly, consider that if $x \in \limsup_{n\to\infty}(A_n \cup B_n)$, then by definition we know that x appears in $\{(A_m \cup B_m)\}_{m\in\mathbb{N}}$ for infinite many times. Observe that either of the cases below must happen:

- x appears in $\{A_m\}_{m\in\mathbb{N}}$ for infinitely many times, then by definition $x \in \limsup_{n\to\infty} A_n$.
- x appears in $\{B_m\}_{m\in\mathbb{N}}$ for infinitely many times, then by definition $x \in \limsup_{n\to\infty} B_n$.

Hence, we get that $x \in \big(\limsup_{n\to\infty} A_n\big) \cup \big(\limsup_{n\to\infty} B_n\big)$, immediately, we finish the proof for

$$\limsup_{n\to\infty}(A_n \cup B_n) \;\subset\; \big(\limsup_{n\to\infty} A_n\big) \cup \big(\limsup_{n\to\infty} B_n\big).$$

Secondly, the other direction is rather trivial, consider that if $x \in \big(\limsup_{n\to\infty} A_n\big) \cup \big(\limsup_{n\to\infty} B_n\big)$, then

$$x \in \limsup_{n\to\infty} A_n \text{ or } x \in \limsup_{n\to\infty} B_n$$

which means that x appears in $\{A_m\}_{m\in\mathbb{N}}$ or $\{B_m\}_{m\in\mathbb{N}}$ for infinitely many times, therefore x appears in $\{(A_m \cup B_m)\}_{m\in\mathbb{N}}$ for infinite many times. This finishes the proof for

$$\limsup_{n\to\infty}(A_n \cup B_n) \;\supset\; \big(\limsup_{n\to\infty} A_n\big) \cup \big(\limsup_{n\to\infty} B_n\big).$$

In conclusion, the limsup of union of two sequences of sets is equivalent to the union of these two limsup sets.

$$\limsup_{n\to\infty}(A_n \cup B_n) \;=\; \big(\limsup_{n\to\infty} A_n\big) \cup \big(\limsup_{n\to\infty} B_n\big).$$

(b) We will argue that the equation (b) is not true by giving a counterexample with corrected statement:

Consider that two sequences of set (A_n) and (B_n) that alternate: one is Ω exactly when the other is $\varnothing$, and they swap roles at each step, and Ω can be any nonempty set, say $\{1\}$:

$$A_n = \begin{cases} \Omega, & n \text{ odd}, \\ \varnothing, & n \text{ even}, \end{cases} \qquad B_n = \begin{cases} \varnothing, & n \text{ odd}, \\ \Omega, & n \text{ even}. \end{cases}$$

Note that $A_n \cap B_n = \varnothing$ for all $n \in \mathbb{N}$, hence $\limsup_{n\to\infty}(A_n \cap B_n) = \varnothing$, while both $\limsup_{n\to\infty} A_n = \limsup_{n\to\infty} B_n = \Omega$ leading to that $\big(\limsup_{n\to\infty} A_n\big) \cap \big(\limsup_{n\to\infty} B_n\big) = \Omega$.

Comment: In general, one can prove that it is a subset but not strictly equivalent:

$$\limsup_{n\to\infty}(A_n \cap B_n) \;\subset\; \big(\limsup_{n\to\infty} A_n\big) \,\cap\, \big(\limsup_{n\to\infty} B_n\big), \tag{b*}$$

since for any $x \in \big(\limsup_{n\to\infty} A_n\big) \cap \big(\limsup_{n\to\infty} B_n\big)$, we only know that there exists two mappings $\{s(i)\}_{i\in\mathbb{N}}$ and $\{t(i)\}_{i\in\mathbb{N}}$ such that x appears in $\{A_{s(i)}\}$ for infinitely many times and also in $\{B_{t(i)}\}$ for infinitely many times, but these sequences of indices might not coincide which generally fails to validate that x appears in $\{A_n \cap B_n\}_{n\in\mathbb{N}}$ for infinite times, as in the counterexample.

(c) We will argue that the equation (c) is not true by using the same counterexample from above:

Now, observe that $A_n \cup B_n = \Omega$ for all $n \in \mathbb{N}$, hence $\liminf_{n\to\infty}(A_n \cup B_n) = \Omega$, while both $\liminf_{n\to\infty} A_n = \liminf_{n\to\infty} B_n = \varnothing$ leading to that $\big(\liminf_{n\to\infty} A_n\big) \cup \big(\liminf_{n\to\infty} B_n\big) = \varnothing$, which contradicts (c) equation.

Comment: In general, one can prove that it is a superset but not strictly equivalent:

$$\liminf_{n\to\infty}(A_n \cup B_n) \;\supset\; \big(\liminf_{n\to\infty} A_n\big) \,\cup\, \big(\liminf_{n\to\infty} B_n\big), \tag{c*}$$

since for any $x \in \liminf_{n\to\infty}(A_n \cup B_n)$, we only know that there exists certain n^* such that eventually x appears in all $\{A_n \cup B_n\}$ for $n \geq n^*$, but this does not distinguish being in $\{A_n\}$ or $\{B_n\}$, which generally fails to validate that x will eventually appears in all $\{A_n\}_{n\in\mathbb{N}}$ but finitely starting ones, neither in $\{B_n\}$.

(d) Consider any $x \in \liminf_{n\to\infty}(A_n \cap B_n)$, by definition, we know there exists certain n^* such that eventually x appears in all $\{A_n \cap B_n\}$ for $n \geq n^*$, immediately, we have that eventually x also appears in both of all the $\{A_n\}$ and $\{B_n\}$ for $n \geq n^*$. By definition of $\liminf$, we get $x \in \liminf_{n\to\infty} A_n$ and $x \in \liminf_{n\to\infty} B_n$. Hence, we proved that

$$\liminf_{n\to\infty}(A_n \cap B_n) \subset \big(\liminf_{n\to\infty} A_n\big) \cap \big(\liminf_{n\to\infty} B_n\big).$$

While for any $x \in \big(\liminf_{n\to\infty} A_n\big) \cap \big(\liminf_{n\to\infty} B_n\big)$, we know both $x \in \liminf_{n\to\infty} A_n$ and $x \in \liminf_{n\to\infty} B_n$ must hold. By definition of $\liminf$ again, there exists $n_1, n_2 \in \mathbb{N}$ such that eventually x appears in $\{A_n\}$ for all $n \geq n_1$ and also in $\{B_n\}$ for all $n \geq n_2$. Now let's take $n^* := \max(n_1, n_2)$, we achieve that for any $n \geq n^*$, x will eventually appears in all $\{A_n \cap B_n\}$ for $n \geq n^*$. This gives the other direction:

$$\liminf_{n\to\infty}(A_n \cap B_n) \supset \big(\liminf_{n\to\infty} A_n\big) \cap \big(\liminf_{n\to\infty} B_n\big).$$

In conclusion, the equation (d) is correct:

$$\liminf_{n\to\infty}(A_n \cap B_n) = \big(\liminf_{n\to\infty} A_n\big) \cap \big(\liminf_{n\to\infty} B_n\big).$$

Lastly, we can now starting to prove that if $A_n \to A$ and $B_n \to B$ (i.e. both limsup and liminf equal), then

$$A_n \cup B_n \to A \cup B \quad \text{as } n \to \infty, \tag{lm1}$$

$$A_n \cap B_n \to A \cap B \quad \text{as } n \to \infty. \tag{lm2}$$

Now since the existence of the limit sets gives that

$$A = \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n \text{ and } B = \limsup_{n\to\infty} B_n = \liminf_{n\to\infty} B_n,$$

(lm1)

Using the inclusion relationship between lim inf and lim sup and the **equation (a)** we derived before, we know

$$\liminf_{n\to\infty}(A_n \cup B_n) \subset \limsup_{n\to\infty}(A_n \cup B_n) = \big(\limsup_{n\to\infty} A_n\big) \cup \big(\limsup_{n\to\infty} B_n\big) = A \cup B$$

In addition, recall that we have the **corrected equation (c^*)** and then we can rewrite $A,\ B$ to get that

$$A \cup B = \big(\liminf_{n\to\infty} A_n\big) \cup \big(\liminf_{n\to\infty} B_n\big) \subset \liminf_{n\to\infty}(A_n \cup B_n)$$

Therefore, we derive that

$$\liminf_{n\to\infty}(A_n \cup B_n) \subset \limsup_{n\to\infty}(A_n \cup B_n) = A \cup B \subset \liminf_{n\to\infty}(A_n \cup B_n)$$

This directly proves that indeed the limit of sets $\{A_n \cup B_n\}$ exists and equal to $(A \cup B)$:

$$\lim_{n\to\infty}(A_n \cup B_n) = A \cup B \ \Big(= \liminf_{n\to\infty}(A_n \cup B_n) = \limsup_{n\to\infty}(A_n \cup B_n)\Big).$$

(lm2)

Similarly, using the inclusion relationship between lim inf and lim sup and the **equation (d)** we derived before,

$$\limsup_{n\to\infty}(A_n \cap B_n) \supset \liminf_{n\to\infty}(A_n \cap B_n) = \big(\liminf_{n\to\infty} A_n\big) \cap \big(\liminf_{n\to\infty} B_n\big) = A \cap B$$

In addition, recall that we have the **corrected equation (b^*)** and then we can rewrite $A,\ B$ to get that

$$A \cap B = \big(\limsup_{n\to\infty} A_n\big) \cap \big(\limsup_{n\to\infty} B_n\big) \supset \limsup_{n\to\infty}(A_n \cap B_n)$$

Therefore, we derive that

$$\limsup_{n\to\infty}(A_n \cap B_n) \supset \liminf_{n\to\infty}(A_n \cap B_n) = A \cap B \supset \limsup_{n\to\infty}(A_n \cap B_n)$$

This directly proves that indeed the limit of sets $\{A_n \cap B_n\}$ exist and equal to $(A \cup B)$:

$$\lim_{n\to\infty}(A_n \cap B_n) = A \cap B \ \Big(= \liminf_{n\to\infty}(A_n \cap B_n) = \limsup_{n\to\infty}(A_n \cap B_n)\Big).$$

$\square$

Definition 1.7 (Limit of Sets)**.** *If* $\liminf A_n = \limsup A_n$*, we say the sequence* (A_n) converges*, and define*

$$\lim_{n\to\infty} A_n := \liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n.$$

Example. Let

$$A_{2n} = \left[\tfrac{1}{n},\, 7 - \tfrac{1}{n}\right], \qquad A_{2n+1} = \left(-\tfrac{1}{n},\, 7 + \tfrac{1}{n}\right).$$

Then

$$\liminf_{n\to\infty} A_n = (0,7), \qquad \limsup_{n\to\infty} A_n = [0,7].$$

Remark. The operations $\liminf$ and $\limsup$ for sets should *not* be confused with $\liminf$ and $\limsup$ for real sequences. For sets, they capture eventual and infinitely-often membership, not numerical bounds.

π-systems, Monotone Classes, and Dynkin Systems

Definition 1.8 (π-system)**.** *A collection* $\mathcal{A}$ *of subsets of* Ω *is a* π-system *if it is closed under finite intersections:*

$$A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}.$$

Definition 1.9 (Monotone Class)**.** *A collection* $\mathcal{M}$ *of subsets of* Ω *is a* monotone class *if it is closed under monotone limits:*

- *If* (A_n) *is an increasing sequence in* $\mathcal{M}$ $(A_n \subseteq A_{n+1})$*, then* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}$*.*
- *If* (A_n) *is a decreasing sequence in* $\mathcal{M}$ $(A_n \supseteq A_{n+1})$*, then* $\bigcap_{n=1}^{\infty} A_n \in \mathcal{M}$*.*

Definition 1.10 (Dynkin System (or λ-system))**.** *A collection* $\mathcal{D}$ *of subsets of* Ω *is a* Dynkin system *if:*

1. $\Omega \in \mathcal{D}$*,*
2. *If* $A, B \in \mathcal{D}$ *with* $A \subseteq B$*, then* $B \setminus A \in \mathcal{D}$*,*
3. *If* (A_n) *are disjoint sets in* $\mathcal{D}$*, then* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{D}$*.*

Remark.A Dynkin system that is also a π-system is automatically a σ-algebra. Indeed, closure under finite intersections and disjoint countable unions together imply closure under countable unions.

Definition 1.11 (Generated σ-algebra)**.** *Let* $\mathcal{A}$ *be a collection of subsets of* Ω*. The* σ-algebra generated by $\mathcal{A}$*, denoted* $\sigma(\mathcal{A})$*, is the smallest* σ*-algebra containing* $\mathcal{A}$*:*

$$\sigma(\mathcal{A}) = \bigcap \{\mathcal{B} : \mathcal{A} \subseteq \mathcal{B},\ \mathcal{B} \text{ is a } \sigma\text{-algebra}\}.$$

Example. Let $\Omega = \{1,2,3,4\}$ and $\mathcal{A} = \{\{1\}\}$. Then $\mathcal{A}$ is not a σ-algebra, but the generated $\sigma(\mathcal{A})$ is

$$\sigma(\mathcal{A}) = \{\{1\}, \{2,3,4\}, \Omega, \emptyset\}.$$

Examples of σ-algebras

Example. Let $\Omega = \{1,2,3,4\}$ and $A = \{\{1\},\{3,4\}\}$. Then the σ-algebra generated by A is

$$\sigma(A) = \big\{\{1\}, \{3,4\}, \{2,3,4\}, \{1,2\}, \{1,3,4\}, \{2\}, \Omega, \emptyset\big\}.$$

Example (Borel σ-algebra)**.**Let $\Omega = \mathbb{R}$ and $A = \{(a,b) : a < b\}$. Then $\sigma(A)$ is called the *Borel σ-algebra*.

Example (Filtration)**.**Let $\Omega = \{X = (x_1, x_2, \dots) : x_i \in \{0,1\}\}$, i.e. the set of infinite sequences of 0's and 1's.

Define $\mathcal{F}_n$ = information after n flips of a coin:

$$\mathcal{F}_1 = \{\{X : x_1 = 1\},\ \{X : x_1 = 0\},\ \Omega,\ \emptyset\},$$

$$\mathcal{F}_2 = \{\{X : x_1 = 1, x_2 = 1\},\ \{X : x_1 = 1, x_2 = 0\},\ \{X : x_1 = 0, x_2 = 1\},\ \{X : x_1 = 0, x_2 = 0\},\ \Omega,\ \emptyset\}.$$

Probability Space Definition

Definition 1.12 (Probability Space). *A* probability space *is a triple* $(\Omega, \mathcal{F}, P)$*, where*

- Ω *is the sample space,*
- $\mathcal{F}$ *is a* σ*-algebra of subsets of* Ω*,*
- $P : \mathcal{F} \to [0, 1]$ *is a probability measure satisfying the Kolmogorov axioms:*
  - *(i)* $P(A) \in [0, 1]$ *for all* $A \in \mathcal{F}$*,*
  - *(ii)* $P(\Omega) = 1$*,*
  - *(iii) (Countable additivity) If* $\{A_i\}$ *are pairwise disjoint, then*

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

Basic Properties of Probability

Remark. 1. If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.

2. $P(\emptyset) = 0$.
3. $P(A^c) = 1 - P(A)$.
4. (Inclusion–Exclusion)

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

   Equivalently,

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

5. If $A_i \in \mathcal{F}$, then

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

Power Set, Algebras of Sets, and Sigma Algebra

Definition 1.13 (Power Set). *For a set* Ω*, the* power set 2^Ω *is the set of all subsets of* Ω*:*

$$2^\Omega = \{A : A \subseteq \Omega\}.$$

Example. If $\Omega = \{1, 2\}$, then

$$2^\Omega = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}.$$

If $\Omega = \{4, 5, 6\}$, then

$$2^\Omega = \{\emptyset, \{4\}, \{5\}, \{6\}, \{4, 5\}, \{4, 6\}, \{5, 6\}, \{4, 5, 6\}\}.$$

This illustrates that if $|\Omega| = n$, then $|2^\Omega| = 2^n$.

Definition 1.14 (Algebra of Sets). *An* algebra $\mathcal{A}$ *on* Ω *is a collection of subsets of* Ω *such that:*

1. $\Omega \in \mathcal{A}$*,*
2. *If* $A \in \mathcal{A}$*, then* $A^c \in \mathcal{A}$*,*
3. *If* $A, B \in \mathcal{A}$*, then* $A \cup B \in \mathcal{A}$*.*

Remark. An algebra is closed under complements and finite unions (and hence finite intersections). It can be viewed as a "mini-universe" of sets stable under the basic operations of set theory.

Definition 1.15 (σ-Algebra). *An algebra* $\mathcal{A}$ *is a* σ-algebra *if it is closed under countable unions:*

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

Definition 1.16 (Borel σ-algebra). *The* Borel σ-algebra $\mathcal{B}(\mathbb{R})$ *is the smallest* σ*-algebra containing all open intervals* $(a, b) \subseteq \mathbb{R}$*.*

Probability Measures

Definition 1.17 (Probability Measure)**.** *A function* $P : \mathcal{A} \to [0, 1]$ *on a* σ*-algebra* $\mathcal{A}$ *is a* probability measure *if*

1. $P(\Omega) = 1$,
2. *(Countable additivity) For disjoint* $A_i \in \mathcal{A}$,

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

Definition 1.18 (Probability Space)**.** *A triple* $(\Omega, \mathcal{A}, P)$ *is a* probability space, *where*

- Ω *is the sample space,*
- $\mathcal{A}$ *is a* σ*-algebra of events,*
- P *is a probability measure.*

Properties of Probability

Proposition 1.1 (Basic Properties)**.** *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space. Then for all* $A, B, A_i \in \mathcal{F}$*:*

(i) $P(\emptyset) = 0, \quad P(A^c) = 1 - P(A)$.

(ii) If $A \subseteq B$*, then* $P(A) \le P(B)$.

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

(iv) (Boole's inequality)

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

Proof. (i) Since $\Omega = A \cup A^c$ (disjoint), we have

$$P(\Omega) = P(A) + P(A^c) = 1,$$

hence $P(A^c) = 1 - P(A)$. Taking $A = \Omega$ gives $P(\emptyset) = 0$.

(ii) If $A \subseteq B$, then $B = A \cup (B \setminus A)$ disjoint. Thus

$$P(B) = P(A) + P(B \setminus A) \ge P(A).$$

(iii) Trick: decompose $A \cup B$ into disjoint parts. Write

$$A \cup B = (A \setminus B) \,\dot\cup\, (B \setminus A) \,\dot\cup\, (A \cap B).$$

Therefore

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

But $P(A) = P(A \setminus B) + P(A \cap B)$ and $P(B) = P(B \setminus A) + P(A \cap B)$. Adding yields the inclusion–exclusion identity:

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

(iv) (Boole's inequality). The spirit of this proof is a *disjointification trick.* Define

$$B_1 = A_1, \qquad B_2 = A_2 \setminus A_1, \qquad B_3 = A_3 \setminus (A_1 \cup A_2), \quad \ldots$$

Then the B_i are disjoint, and

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i.$$

Hence by countable additivity,

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

Since $B_i \subseteq A_i$, we have $P(B_i) \le P(A_i)$. Thus

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

□

Remark (Spirit of the Trick)**.**is

- The inclusion–exclusion formula relies on *splitting overlapping sets into disjoint pieces.* This avoids over-counting.
- Boole's inequality relies on constructing a *disjoint cover* B_i of $\cup_i A_i$ by "peeling off" previously counted parts. This is a standard probability trick: make things disjoint to apply additivity.

Theorem 1.4 (Continuity from Below and Above)**.** *Let $\{A_n\}$ be a monotone sequence of sets.*

1. *If $A_1 \subseteq A_2 \subseteq \dots$, then*
$$\lim_{n\to\infty} P(A_n) = P\Big(\bigcup_{n=1}^{\infty} A_n\Big).$$
2. *If $A_1 \supseteq A_2 \supseteq \dots$, then*
$$\lim_{n\to\infty} P(A_n) = P\Big(\bigcap_{n=1}^{\infty} A_n\Big).$$

Proof Idea. For (i), define disjoint increments $B_1 = A_1$, $B_{n+1} = A_{n+1} \setminus A_n$. Then $A = \cup_n A_n = \cup_n B_n$ disjoint, and

$$P(A_n) = \sum_{k=1}^{n} P(B_k), \quad P(A) = \sum_{k=1}^{\infty} P(B_k).$$

Thus $\lim_n P(A_n) = P(A)$. For (ii), apply the result to complements: A_n^c is increasing, and use $P(A) = 1 - P(A^c)$. □

Proposition 1.2 (Liminf and Limsup Inequalities)**.** *For any sequence $\{A_n\} \subseteq \mathcal{F}$,*

$$P(\liminf A_n) \le \liminf_{n\to\infty} P(A_n) \le \limsup_{n\to\infty} P(A_n) \le P(\limsup A_n).$$

If $\lim A_n$ exists (i.e. $\liminf A_n = \limsup A_n$), then

$$\lim_{n\to\infty} P(A_n) = P(\lim A_n).$$

Spirit. This result captures the *compatibility of limits and probability.* - The trick: rewrite $\liminf A_n$ and $\limsup A_n$ using $\cap\cup$ and $\cup\cap$, then apply continuity from below/above. - The inequality arises because $\liminf A_n \subseteq$ "eventually in A_n" $\subseteq \limsup A_n$. □

Countable vs. Uncountable Sets in Probability Spaces

Example (Countable Probability Space)**.**Let $\Omega = \mathbb{N}$, $\mathcal{F} = 2^{\mathbb{N}}$, and define

$$P(\{k\}) = 2^{-k}, \qquad k \in \mathbb{N}.$$

Since $\sum_{k=1}^{\infty} 2^{-k} = 1$, this defines a probability measure. For instance, the probability of choosing an even number is

$$P(\{\text{even}\}) = \sum_{n=1}^{\infty} P(\{2n\}) = \sum_{n=1}^{\infty} 2^{-2n} = \frac{1}{4} \cdot \frac{1}{1-\frac{1}{4}} = \tfrac{1}{3}.$$

Example.(Lebesgue Measure on $[0,1]$) Let $\Omega = [0,1]$, $\mathcal{F}$ the Borel σ-algebra, and $P((a,b)) = b - a$. Then for any point $x \in [0,1]$,

$$P(\{x\}) = \lim_{n\to\infty} P\big((x - \tfrac{1}{n},\, x + \tfrac{1}{n}) \cap [0,1]\big) = 0.$$

Thus every singleton has measure zero.

More generally, if $A = \{a_n\}_{n=1}^{\infty} \subseteq [0,1]$ is countable, then

$$P(A) = \sum_{n=1}^{\infty} P(\{a_n\}) = 0.$$

In particular,

$$P(\mathbb{Q} \cap [0,1]) = 0.$$

Remark (Countable vs. Uncountable Additivity)**.**While countable sets in $[0,1]$ always have probability zero, the uncountable union

$$[0,1] = \bigcup_{x\in[0,1]} \{x\}$$

satisfies

$$P([0,1]) = 1 \quad \neq \quad \sum_{x\in[0,1]} P(\{x\}) = 0.$$

This illustrates why probability measures are *countably additive*, but not uncountably additive.

Conditioning and Independence

Conditional Probability

Definition 1.19 (Conditional Probability)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space. For $A, B \in \mathcal{F}$ with $P(B) > 0$, the* conditional probability *of A given B is*

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

Independence of Events

Definition 1.20 (Independence)**.** *Two events $A, B \in \mathcal{F}$ are* independent *if*

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

equivalently, $P(A \mid B) = P(A)$ (when $P(B) > 0$) and $P(B \mid A) = P(B)$ (when $P(A) > 0$).

Remark. If $P(A) = 0$ or $P(A) = 1$, then A is independent of every $B \in \mathcal{F}$.

Proposition 1.3 (Closure Properties)**.** *If A and B are independent, then so are:*

$$A \text{ and } B^c, \qquad A^c \text{ and } B, \qquad A^c \text{ and } B^c.$$

Proof. Suppose $P(A \cap B) = P(A)P(B)$. Then

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c).$$

Similarly for the other cases. □

Independence of Families of Events

Definition 1.21 (Mutual Independence)**.** *A collection $\{A_k : 1 \le k \le n\}$ is* independent *if*

$$P\left(\bigcap_{k=1}^{m} A_{i_k}\right) = \prod_{k=1}^{m} P(A_{i_k}), \quad \forall\ 1 \le i_1 < i_2 < \cdots < i_m \le n.$$

Definition 1.22 (Pairwise Independence)**.** *A collection $\{A_k : 1 \le k \le n\}$ is* pairwise independent *if*

$$P(A_i \cap A_j) = P(A_i)P(A_j), \quad \forall i \neq j.$$

Remark. Mutual independence $\implies$ pairwise independence, but the converse is not true. Pairwise independence is strictly weaker.

Example (Pairwise Independent but not Mutually Independent)**.** Consider two fair coin flips with sample space $\Omega = \{00, 01, 10, 11\}$ and uniform probability $P(x) = 1/4$. Define events:

$$A = \{01, 11\} \quad (\text{“2nd coin is head”}), \quad B = \{10, 11\} \quad (\text{“1st coin is head”}), \quad C = \{01, 10\} \quad (\text{“exactly one head”}).$$

We check:

$$P(A \cap B) = P(\{11\}) = \tfrac{1}{4} = P(A)P(B), \quad P(A \cap C) = \tfrac{1}{4} = P(A)P(C), \quad P(B \cap C) = \tfrac{1}{4} = P(B)P(C).$$

So (A, B, C) are pairwise independent. However,

$$P(A \cap B \cap C) = 0 \neq \tfrac{1}{4} = P(A)P(B)P(C),$$

so they are not mutually independent.

Definition 1.23 (Countable Independence)**.** *A countable family $\{A_k : k \geq 1\}$ is independent if every finite subfamily is independent.*

Law of Total Probability

Definition 1.24 (Partition)**.** *A collection $\{H_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ is a* partition *of Ω if*

$$H_i \cap H_j = \emptyset \;\; (i \neq j), \qquad \bigcup_{i=1}^{\infty} H_i = \Omega.$$

Theorem 1.5 (Law of Total Probability)**.** *If $\{H_i\}_{i=1}^{\infty}$ is a partition with $P(H_i) > 0$, then for any $A \in \mathcal{F}$,*

$$P(A) = \sum_{i=1}^{\infty} P(A \mid H_i)P(H_i).$$

Proof. We can write $A = \bigcup_{i=1}^{\infty}(A \cap H_i)$, a disjoint union. Then

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

□

Example. Roll a fair die, and then flip a coin as many times as the die shows. Let $A =$ “total number of heads is 3.” Partition by die outcome: $H_i = \{\text{die shows } i\}$.

$$P(A \mid H_i) = \begin{cases} 0, & i < 3, \\ \binom{i}{3}\left(\frac{1}{2}\right)^i, & i \geq 3. \end{cases}$$

Since $P(H_i) = 1/6$,

$$P(A) = \sum_{i=3}^{6} \binom{i}{3} \left(\tfrac{1}{2}\right)^i \cdot \tfrac{1}{6}.$$

Remark (Spirit of the Law)**.** The law of total probability decomposes a complicated event A into simpler conditional pieces along a partition $\{H_i\}$. It is the foundation for *Bayes' Theorem* and for reasoning under uncertainty: probabilities are consistent across different levels of information.

2 Unit 2: Random Variables

Measurable Functions and Random Elements

Definition 2.1 (Random Element)**.** *Let* $(\Omega, \mathcal{F})$ *and* $(S, \mathcal{S})$ *be measurable spaces. A mapping* $X : \Omega \to S$ *is called a* random element*. If* $S = \mathbb{R}$ *with* $\mathcal{S} = \mathcal{B}(\mathbb{R})$*, then* X *is a* random variable*.*

Definition 2.2 (Measurable Function)**.** $X : \Omega \to \mathbb{R}$ *is measurable if*

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F} \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

Example (Indicator Function)**.**If $A \in \mathcal{F}$, then

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

is a random variable.

Proposition 2.1 (Closure)**.** *If* $X_1, \dots, X_n$ *are random variables and* $g : \mathbb{R}^n \to \mathbb{R}$ *is Borel measurable, then* $g(X_1, \dots, X_n)$ *is a random variable.*

Proposition 2.2 (Limits)**.** *If* $X_n \to X$ *almost surely and each* X_n *is a random variable, then* X *is a random variable.*

Equivalent Random Variables and Measurability

Definition 2.3 (Equivalent Random Variables)**.** *Two random variables* $X, Y : \Omega \to \mathbb{R}$ *are called* equivalent *if*

$$P(X = Y) = 1 \quad \textit{equivalently,} \quad P(X \neq Y) = 0.$$

In this case, X *and* Y *are indistinguishable under* P*.*

Definition 2.4 (Distribution of a Random Variable)**.** *Any measurable random variable* $X : (\Omega, \mathcal{F}, P) \to \mathbb{R}$ *induces a probability measure* μ_X *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *by*

$$\mu_X(A) = P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}), \quad A \in \mathcal{B}(\mathbb{R}).$$

This μ_X *is called the* distribution of X*.*

Definition 2.5 (Equal in Distribution)**.** *Two random variables* X *and* Y *(possibly on different probability spaces) are said to be* equal in distribution*, written* $X \overset{d}{=} Y$*, if they induce the same law:*

$$\mu_X = \mu_Y.$$

Example. Let $\Omega = (0, 1]$ with Lebesgue measure. Define

$$X = 1_{(0,\frac{1}{2}]}, \qquad Y = 1_{(\frac{1}{2},1]}.$$

Then X and Y do not satisfy $P(X = Y) = 1$ (in fact, $P(X = Y) = 0$). However, both are Bernoulli(1/2) random variables, hence $X \overset{d}{=} Y$.

Theorem 2.1 (Function Composition)**.** *If* $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is measurable and* $g : \mathbb{R} \to \mathbb{R}$ *is Borel-measurable, then the composition*

$$g \circ X : \Omega \to \mathbb{R}$$

is a random variable.

Proof. Let $A \in \mathcal{B}(\mathbb{R})$. Then

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

since $g^{-1}(A) \in \mathcal{B}(\mathbb{R})$ and X is measurable. Hence $g(X)$ is measurable. ☐

Remark. This shows that applying measurable transformations (e.g. sin, exp, log, polynomials) to random variables always yields another random variable.

Proposition 2.3 (Algebra of Measurable Functions)**.** *Let X, Y be random variables. Then:*

- *$X + Y$, $X - Y$, and cX (for $c \in \mathbb{R}$) are random variables.*
- *XY is a random variable.*
- *$\max\{X, Y\}$ and $\min\{X, Y\}$ are random variables.*

Sketch. For sums: $\{X+Y < t\}$ can be expressed as a countable union of intersections involving $\{X < q\}, \{Y < t - q\}$ with $q \in \mathbb{Q}$, hence measurable. For products: express $\{XY < t\}$ via rational bounds depending on the signs of X, Y. For max/min: note that

$$\max\{X, Y\} = \tfrac{1}{2}(X + Y + |X - Y|), \quad \min\{X, Y\} = \tfrac{1}{2}(X + Y - |X - Y|).$$

All are measurable as they are compositions of measurable maps. □

Theorem 2.2 (Sup/Inf of Random Variables)**.** *Let $\{X_n\}_{n\geq 1}$ be random variables. Then*

$$\sup_n X_n, \quad \inf_n X_n, \quad \limsup_{n\to\infty} X_n, \quad \liminf_{n\to\infty} X_n$$

are all random variables.

Proof. For $\sup_n X_n$:

$$\{\sup_n X_n \leq t\} = \bigcap_{n=1}^{\infty} \{X_n \leq t\},$$

which is measurable as a countable intersection of measurable sets. Similarly for $\inf_n$. For lim sup:

$$\{\limsup_{n\to\infty} X_n \leq t\} = \bigcap_{m=1}^{\infty} \bigcup_{n\geq m} \{X_n \leq t\},$$

which is measurable. Analogous for lim inf. □

Remark. The constructions of sup, inf, lim sup, and lim inf are all "pointwise" operations, hence measurability follows from closure of $\mathcal{F}$ under countable unions/intersections. This highlights an important principle: *measurability is stable under pointwise limits and algebraic operations.*

Limsup, Liminf, and Measurability of Limits

Definition 2.6 (Limsup and Liminf of Numerical Sequences)**.** *Let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence of real numbers. We define*

$$\limsup_{n\to\infty} a_n = \inf_{m\geq 1} \sup_{n\geq m} a_n, \qquad \liminf_{n\to\infty} a_n = \sup_{m\geq 1} \inf_{n\geq m} a_n.$$

Remark. Intuitively:

- lim sup tracks the "eventual upper envelope" of the sequence — the smallest ceiling that still contains infinitely many terms.
- lim inf tracks the "eventual lower envelope" — the largest floor that still contains infinitely many terms.

It follows that $\liminf a_n \leq \limsup a_n$ always, and if they coincide, the common value is the limit of the sequence.

Theorem 2.3 (Measurability of lim sup and lim inf)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of measurable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then*

$$\limsup_{n\to\infty} X_n(\omega), \qquad \liminf_{n\to\infty} X_n(\omega)$$

are measurable random variables.

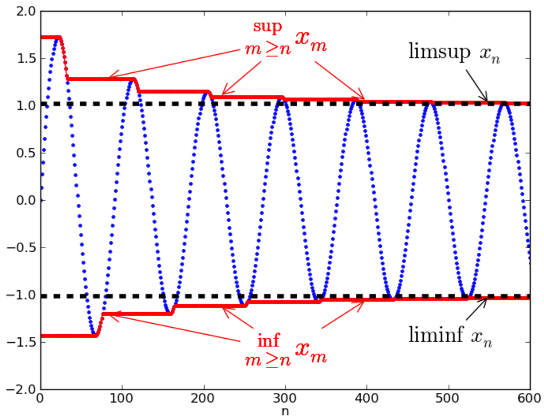

Figure 1: An illustration of limit superior and limit inferior.

Proof. Fix $\omega \in \Omega$. Define

$$Y_m(\omega) = \sup_{n \geq m} X_n(\omega).$$

Each Y_m is measurable because measurability is preserved under sup of a countable family. Then

$$\limsup_{n\to\infty} X_n(\omega) = \inf_{m\geq 1} Y_m(\omega),$$

which is measurable as an infimum of measurable functions. The case of lim inf follows similarly. □

Remark (Alternative Characterization)**.**A useful identity is:

$$\{\limsup_{n\to\infty} X_n \leq x\} = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{X_n \leq x\}.$$

This set belongs to $\mathcal{F}$ because it is built from countable unions and intersections of measurable sets. This alternative description is often more practical when proving measurability.

Theorem 2.4 (Measurability of Limits When They Exist)**.** *Let $\{X_n\}$ be measurable random variables. Suppose*

$$\lim_{n\to\infty} X_n(\omega)$$

exists for all ω in a set of probability one (i.e., $\limsup X_n = \liminf X_n$ almost surely). Then

$$\lim_{n\to\infty} X_n$$

is a measurable random variable.

Remark (Spirit of the Result)**.**This theorem emphasizes an important principle: the pointwise limit of measurable random variables is again measurable (at least almost surely). This is essential for probability theory, since random variables are often defined as limits of simpler approximations (e.g., in the construction of expectations, martingales, or stochastic processes).

The deeper idea is that measurability is robust under limits — which is why probability theory works so well with infinite sequences and limiting arguments.

Sigma-Algebra Generated by a Random Variable

Definition 2.7 (Sigma-Algebra Generated by a Random Variable)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ a measurable random variable. The* sigma-algebra generated by X*, denoted $\sigma(X)$, is defined by*

$$\sigma(X) = \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R})\},$$

where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra on $\mathbb{R}$. Equivalently, $\sigma(X)$ is the smallest σ-algebra with respect to which X is measurable.

Example (Flipping Two Coins)**.**Consider $\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$ with $\mathcal{F} = 2^\Omega$ and $X(\omega)$ defined as the total number of heads:

$$X(0,0) = 0, \quad X(0,1) = X(1,0) = 1, \quad X(1,1) = 2.$$

Then

$$\sigma(X) = \big\{\emptyset,\ \Omega,\ \{(0,0)\},\ \{(1,1)\},\ \{(0,1),(1,0)\},\ \{(0,0),(0,1),(1,0)\},\ \{(0,1),(1,0),(1,1)\}\big\}.$$

Remark (Interpretation)**.**Two sample points $\omega_1, \omega_2 \in \Omega$ are *indistinguishable with respect to X* if $X(\omega_1) = X(\omega_2)$. In that case, for every $A \in \sigma(X)$, either both ω_1, ω_2 belong to A or neither does. This captures the idea that $\sigma(X)$ contains *exactly the information that can be revealed by knowing the value of X*. It "forgets" everything else about the underlying ω.

Distribution and Distribution Functions

Definition 2.8 (Distribution of a Random Variable)**.** *The* distribution *of a random variable $X : \Omega \to \mathbb{R}$ is the probability measure μ_X on $\mathcal{B}(\mathbb{R})$ defined by*

$$\mu_X(A) = \mathbb{P}(X \in A), \qquad \forall A \in \mathcal{B}(\mathbb{R}).$$

Definition 2.9 (Distribution Function)**.** *The* distribution function $F_X : \mathbb{R} \to [0,1]$ *of X is given by*

$$F_X(x) = \mu_X((-\infty, x]) = \mathbb{P}(X \le x).$$

Remark (Properties of Distribution Functions)**.**The distribution function F_X satisfies:

- F_X is non-decreasing.
- F_X is right-continuous: $\lim_{\varepsilon \downarrow 0} F_X(x+\varepsilon) = F_X(x)$.
- $\lim_{x\to-\infty} F_X(x) = 0,\ \lim_{x\to+\infty} F_X(x) = 1$.
- The jump of F_X at x equals the probability mass at x:

$$\mathbb{P}(X = x) = F_X(x) - \lim_{\varepsilon\downarrow 0} F_X(x-\varepsilon).$$

Remark (Interpretation)**.**The distribution μ_X "pushes forward" the probability measure $\mathbb{P}$ from Ω to $\mathbb{R}$. In this sense, μ_X tells us *everything probabilistic about X*, independently of the underlying sample space.

The function F_X is simply the cumulative form of this measure. Right-continuity arises because $(-\infty, x]$ are the basic generators of the Borel σ-algebra, and monotonicity comes from set inclusion. The jumps of F_X correspond to atoms of the distribution — this explains why discrete distributions appear as step functions.

Types of Random Variables

Definition 2.10 (Simple Random Variable)**.** $X(\omega) = \sum_{k=1}^n x_k 1_{A_k}(\omega)$ *with $x_k \in \mathbb{R}$, $A_k \in \mathcal{F}$, $\{A_k\}$ a measurable partition of Ω.*

Definition 2.11 (Equivalent Random Variables)**.** *X and Y are* equivalent *if $P(X = Y) = 1$, i.e. they differ only on a null set. Notation: $X \overset{a.s.}{=} Y$.*

Lemma 2.1 (Approximation Lemma)**.** *Every nonnegative random variable X can be approximated by an increasing sequence of simple random variables $X_n \uparrow X$ pointwise.*

Distribution of a Random Variable

Definition 2.12 (Induced Measure)**.** *Each random variable $X : \Omega \to \mathbb{R}$ induces a probability measure μ_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by*

$$\mu_X(A) = P(X \in A) = P(\{\omega : X(\omega) \in A\}).$$

Definition 2.13 (Distribution Function)**.** *The cumulative distribution function (CDF) of X is*

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

Proposition 2.4 (Properties of CDF)**.** *1. F_X is non-decreasing.*

2. F_X is right-continuous.

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

4. $P(a < X \le b) = F_X(b) - F_X(a)$.

5. $P(X = x) = F_X(x) - \lim_{y\uparrow x} F_X(y)$.

Remark. Every probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ corresponds uniquely to a distribution function, and vice versa.

Definition 2.14 (Equality in Distribution)**.** *$X \overset{d}{=} Y$ if $F_X = F_Y$ (equivalently $\mu_X = \mu_Y$), even if X, Y live on different spaces.*

4 Types of Distributions

- **Discrete:** $P(X = x_k) > 0$ on a countable set $\{x_k\}$.
- **Continuous:** $P(X = x) = 0$ for all x.
- **Absolutely continuous:** $F_X(x) = \int_{-\infty}^{x} f(t)\,dt$ for a density f.
- **Singular continuous:** F continuous, $F'(x) = 0$ a.e., e.g. Cantor distribution.

Cantor Function as Counterexample

Example (Cantor Function)**.**The Cantor function $C : [0,1] \to [0,1]$ is continuous, non-decreasing, and satisfies

$$C(0) = 0, \qquad C(1) = 1.$$

It has the following properties:

- $C'(x) = 0$ for almost every $x \in [0,1]$ (with respect to Lebesgue measure).
- C is not constant; in fact, it strictly increases along the Cantor set.
- C is not absolutely continuous, since

  $$C(x) - C(0) \ne \int_0^x C'(t)\,dt,$$

  the right-hand side being identically zero.

Therefore the Cantor function illustrates that *continuity and differentiability a.e. are not enough* to guarantee absolute continuity. Its associated measure is the **Cantor distribution**, which is a prime example of a *singular continuous measure.*

Absolute Continuity and The Fundamental Theorem of Calculus

Theorem 2.5 (Fundamental Theorem of Calculus for Lebesgue Integrals [Folland Thm.3.35)**.** *Let $-\infty < a < b < \infty$ and $F : [a,b] \to \mathbb{C}$. The following are equivalent:*

(a) F is absolutely continuous on $[a,b]$.

(b) There exists $f \in L^1([a,b])$ such that

$$F(x) - F(a) = \int_a^x f(t)\,dt, \qquad \forall x \in [a,b].$$

(c) F is differentiable almost everywhere on $[a,b]$, with $F' \in L^1([a,b])$, and

$$F(x) - F(a) = \int_a^x F'(t)\,dt, \qquad \forall x \in [a,b].$$

Remark (Interpretation)**.**This theorem characterizes absolute continuity in three equivalent ways:

- **Definition-based:** F is absolutely continuous if small total length of intervals implies small total variation of F.
- **Integral form:** Absolutely continuous functions are precisely those which can be written as indefinite Lebesgue integrals of some L^1 function.
- **Derivative form:** They are differentiable almost everywhere, their derivative belongs to L^1, and the classical fundamental theorem of calculus holds in the Lebesgue sense.

Thus, absolute continuity identifies the "good" class of functions where differentiation and integration are perfectly compatible.

Common Distributions

Example (Discrete)**.**
- Bernoulli(p): $P(X = 1) = p$, $P(X = 0) = 1 - p$.
- Binomial(n, p): $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$.
- Geometric(p): $P(X = k) = (1 - p)^{k-1} p$, $k \geq 1$.
- Poisson(λ): $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$.

Example (Absolutely Continuous)**.**
- Uniform(a, b): $f(x) = \frac{1}{b-a}$ on (a, b).
- Exponential(λ): $f(x) = \lambda e^{-\lambda x}$ on $[0, \infty)$.
- Normal(μ, σ^2): $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$.

Remark (Memoryless Property)**.**Among all discrete distributions, the *Geometric* distribution is the only one with the memoryless property:

$$P(X > m + n \mid X > m) = P(X > n).$$

Among continuous distributions, the unique memoryless law is the *Exponential* distribution:

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

Remark. Convolutions describe sums of independent random variables:

$$F_{X+Y} = F_X * F_Y, \qquad f_{X+Y} = f_X * f_Y.$$

If one variable has a density, then the sum does as well.

Random Vectors and Joint Distributions

Definition 2.15 (Random Vector)**.** *An n-dimensional* random vector *is*

$$X = (X_1, X_2, \ldots, X_n)^\top,$$

where each X_k is a random variable.

Definition 2.16 (Joint Distribution Function)**.** *For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, the joint distribution of X is*

$$F_X(x) = P(X \leq x) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

- **Discrete case:** Characterized by the joint probability function

$$p_X(x) = P(X = x).$$

- **Absolutely continuous case:** Characterized by the joint density

$$f_X(x) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_X(x).$$

Definition 2.17 (Marginal Distribution). *If $X = (X, Y)$ with joint density $f_{X,Y}$, the marginal density of X is*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy.$$

Remark. Marginals describe lower-dimensional components of a joint distribution, but they do not in general determine the joint distribution uniquely.

Random Elements

Definition 2.18 (Random Element). *Let $(\Omega, \mathcal{F})$ be a measurable space and $(S, \mathcal{S})$ a measurable metric space. A* random element *is a measurable map*

$$X : \Omega \to S,$$

meaning that

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

Remark. Note that

- Random variables and random vectors are special cases of random elements with $S = \mathbb{R}$ or $S = \mathbb{R}^n$.
- The definition of measurability is the same as for ordinary random variables: the preimage of a measurable set is an event.
- The generality lies in allowing arbitrary measurable spaces $(S, \mathcal{S})$ as state spaces.

Borel–Cantelli Lemmas

Theorem 2.6 (First Borel–Cantelli). *Let $\{A_n\}$ be a sequence of events. If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then*

$$P(A_n \text{ i.o.}) = P\Big(\limsup_{n\to\infty} A_n\Big) = 0.$$

Proof Sketch. If $\sum P(A_n) < \infty$, then by the convergence of the series, $P(\bigcup_{n \geq N} A_n) \to 0$ as $N \to \infty$. But $\limsup A_n = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} A_n$, so its probability is zero. □

Theorem 2.7 (Second Borel–Cantelli). *If the A_n are independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$, then*

$$P(A_n \text{ i.o.}) = 1.$$

Proof Sketch. Independence allows one to show $P(\bigcap_{n \geq N} A_n^c) = 0$ for all N. Hence the complement of $\limsup A_n$ has probability zero. □

Expectation

Expectation is defined as the probabilistic version of integration. We construct it step by step:

Step 1: Simple random variables.

Definition 2.19 (Simple function). *Let $\{A_i\}_{i=1}^n$ be disjoint measurable sets and $c_i \in \mathbb{R}$. A* simple random variable *is a function*

$$X(\omega) = \sum_{i=1}^{n} c_i \mathbf{1}_{A_i}(\omega), \qquad A_i \cap A_j = \emptyset \ (i \neq j).$$

The expectation is defined as

$$\mathbb{E}[X] = \sum_{i=1}^{n} c_i\, \mathbb{P}(A_i).$$

This can be interpreted as a weighted average of the values c_i with weights $\mathbb{P}(A_i)$.

Step 2: Nonnegative random variables.

Definition 2.20 (Nonnegative random variables)**.** *If $X(\omega) \geq 0$, then*

$$\mathbb{E}[X] = \sup\{\mathbb{E}[Y] : 0 \leq Y \leq X,\ Y \text{ simple}\}.$$

Here, $\mathbb{E}[X]$ is allowed to be $+\infty$.

Approximation trick: Let

$$X_n(\omega) = \begin{cases} \frac{k-1}{2^n}, & \frac{k-1}{2^n} \leq X(\omega) < \frac{k}{2^n},\ k = 1, \ldots, n2^n, \\ n, & X(\omega) \geq n. \end{cases}$$

Then $X_n \uparrow X$ as $n \to \infty$, and

$$\mathbb{E}[X] = \lim_{n\to\infty} \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{P}\big(\tfrac{k-1}{2^n} \leq X < \tfrac{k}{2^n}\big).$$

This provides a concrete approximation scheme for expectations of nonnegative random variables.

Step 3: General random variables.

Definition 2.21. *Let $X : \Omega \to \mathbb{R}$ be a measurable random variable. Define the positive and negative parts of X as*

$$X^+ := \max(X, 0) = |X| \cdot \mathbf{1}_{\{X \geq 0\}}, \qquad X^- := \max(-X, 0) = |X| \cdot \mathbf{1}_{\{X<0\}}.$$

Clearly, $X = X^+ - X^-$ and $|X| = X^+ + X^-$.

- *If $\mathbb{E}[X^+] < \infty$ and $\mathbb{E}[X^-] < \infty$, then we define*

  $$\mathbb{E}[X] := \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

  In this case, we say that X is integrable.

- *If exactly one of $\mathbb{E}[X^+], \mathbb{E}[X^-]$ is infinite and the other is finite, then we set $\mathbb{E}[X] = \pm\infty$ accordingly.*

- *If both $\mathbb{E}[X^+] = \mathbb{E}[X^-] = \infty$, then $\mathbb{E}[X]$ is undefined.*

Properties and Convergence of Expectation

Basic Properties of Expectation

Let X, Y be random variables and $c \in \mathbb{R}$. Then:

- If $\mathbb{P}(X = 0) = 1$, then $\mathbb{E}[X] = 0$.
- If $\mathbb{P}(X \geq 0) = 1$, then $\mathbb{E}[X] \geq 0$.
- If $\mathbb{E}[X] = 0$ and $X \geq 0$, then $\mathbb{P}(X = 0) = 1$.
- Linearity: $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, and $\mathbb{E}[cX] = c\mathbb{E}[X]$.
- Monotonicity: If $X \leq Y$ almost surely, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.
- If $\mathbb{P}(X = Y) = 1$, then $\mathbb{E}[X] = \mathbb{E}[Y]$.
- If $\mathbb{E}[|X - Y|] = 0$, then $\mathbb{P}(X = Y) = 1$.

Convergence of Expectation

A natural question: suppose $X_n \to X$ almost surely, i.e.

$$\mathbb{P}\Big(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\Big) = 1.$$

Do we necessarily have

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X] \ ?$$

Answer: No, not in general.

Example (Counterexample: "Skinny Tall Rectangles")**.**Let $\Omega = (0,1)$, $\mathcal{F}$ the Borel σ-algebra, and $\mathbb{P}$ the Lebesgue measure. Define

$$X_n(\omega) = \begin{cases} n, & \omega \in (0, \frac{1}{n}], \\ 0, & \omega \in (\frac{1}{n}, 1]. \end{cases}$$

Then for each n,

$$\mathbb{E}[X_n] = n \cdot \tfrac{1}{n} + 0 \cdot \Big(1 - \tfrac{1}{n}\Big) = 1.$$

On the other hand, $X_n(\omega) \to 0$ for almost every $\omega \in (0,1)$, so $X = 0$ a.s. and hence $\mathbb{E}[X] = 0$.

Thus

$$\lim_{n\to\infty} \mathbb{E}[X_n] = 1 \quad \neq \quad \mathbb{E}[X] = 0.$$

Remark. This shows that almost sure convergence does not in general imply convergence of expectations. Additional conditions (such as *dominated convergence theorem* or *monotone convergence theorem*) are required to interchange limits and expectation.

Bounded Convergence Theorem

Theorem 2.8 (Bounded Convergence Theorem)**.** *Suppose $X_n \to X$ almost surely, and there exists a constant $M > 0$ such that*

$$|X_n(\omega)| \leq M \quad \textit{for all } n \textit{ and all } \omega.$$

Then

$$\lim_{n\to\infty} \mathbb{E}[X_n] \ = \ \mathbb{E}[X].$$

Proof sketch. We want to show $\mathbb{E}[X_n - X] \to 0$.

Fix $\delta > 0$. Note that for each ω,

$$-\delta\, \mathbf{1}_{\{|X_n - X| < \delta\}} - 2M\, \mathbf{1}_{\{|X_n - X| \geq \delta\}} \ \leq \ X_n - X \ \leq \ \delta\, \mathbf{1}_{\{|X_n - X| < \delta\}} + 2M\, \mathbf{1}_{\{|X_n - X| \geq \delta\}}.$$

Taking expectations and applying monotonicity,

$$-2M\, \mathbb{P}(|X_n - X| \geq \delta) - \delta\mathbb{P}(|X_n - X| < \delta) \ \leq \ \mathbb{E}[X_n - X] \ \leq \ 2M\, \mathbb{P}(|X_n - X| \geq \delta) + \delta.$$

Since $X_n \to X$ almost surely, we have $\mathbb{P}(|X_n - X| \geq \delta) \to 0$. Thus letting $n \to \infty$,

$$-\delta \ \leq \ \liminf_{n\to\infty} \mathbb{E}[X_n - X] \ \leq \ \limsup_{n\to\infty} \mathbb{E}[X_n - X] \ \leq \ \delta.$$

Because $\delta > 0$ was arbitrary, $\mathbb{E}[X_n - X] \to 0$, i.e. $\mathbb{E}[X_n] \to \mathbb{E}[X]$. □

Remark (Trick:)**.**The key idea is to control $X_n - X$ using two bounds simultaneously:

- A *local bound* ($|X_n - X| < \delta$) gives a small error δ.
- A *global bound* ($|X_n - X| \geq \delta$) uses the uniform bound M and the fact that $\mathbb{P}(|X_n - X| \geq \delta) \to 0$.

By combining these via indicator functions and taking expectations, we get a bound of the form

$$|\mathbb{E}[X_n - X]| \leq 2M\, \mathbb{P}(|X_n - X| \geq \delta) + \delta,$$

which vanishes in the limit. This is a recurring theme in measure-theoretic convergence proofs.

Properties of Expectation

Proposition 2.5 (Basic Properties)**.** *For random variables X, Y and constants a, b:*

1. ***Linearity:*** $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$.
2. ***Monotonicity:*** *If $X \geq 0$ a.s., then* $\mathbb{E}[X] \geq 0$.
3. ***Triangle inequality:*** $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

Theorem 2.9 (Law of the Unconscious Statistician (LOTUS))**.** *If X has distribution μ_X and $g : \mathbb{R} \to \mathbb{R}$ is measurable, then*

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)\, \mu_X(dx).$$

In particular,

- *If X has density f_X:* $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$.
- *If X is discrete:* $\mathbb{E}[g(X)] = \sum_x g(x) P(X = x)$.

Convergence Theorems

Theorem 2.10 (Fatou's Lemma)**.** *Let $\{X_n\}_{n\geq 1}$ be a sequence of nonnegative random variables. Then*

$$\mathbb{E}\Big[\liminf_{n\to\infty} X_n\Big] \;\leq\; \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Proof. The key idea is to compare $\liminf X_n$ with a sequence of simple random variables that approximate it from below, so that we can apply bounded convergence.

Step 1. For any bounded simple random variable Y such that

$$0 \leq Y(\omega) \leq \liminf_{n\to\infty} X_n(\omega),$$

we want to show

$$\mathbb{E}[Y] \;\leq\; \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Step 2. Define for each n,

$$Y_n(\omega) := \min\{\, Y(\omega), X_n(\omega)\,\}.$$

This is the "best of both worlds" construction: it guarantees

$$0 \leq Y_n(\omega) \leq Y(\omega) \quad \text{and} \quad Y_n(\omega) \leq X_n(\omega).$$

Thus, Y_n is bounded by Y, but also never exceeds X_n.

Step 3. Now define

$$Z_n(\omega) := \inf_{m\geq n} Y_m(\omega).$$

Then (Z_n) is an increasing sequence (since the index set shrinks as n grows), and by construction,

$$\lim_{n\to\infty} Z_n(\omega) = \liminf_{n\to\infty} Y_n(\omega) \;\leq\; \liminf_{n\to\infty} X_n(\omega).$$

Step 4. By the monotone convergence theorem (applied to the nonnegative increasing sequence Z_n),

$$\mathbb{E}\Big[\lim_{n\to\infty} Z_n\Big] = \lim_{n\to\infty} \mathbb{E}[Z_n].$$

But since $Z_n \leq Y_n \leq X_n$, we get

$$\mathbb{E}[Z_n] \leq \mathbb{E}[Y_n] \leq \mathbb{E}[X_n].$$

Step 5. Combining these,

$$\mathbb{E}[Y] \;\leq\; \lim_{n\to\infty} \mathbb{E}[Z_n] \;\leq\; \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Step 6. Finally, take the supremum over all such bounded simple Y with $Y \leq \liminf X_n$. By the definition of the Lebesgue integral, this supremum equals $\mathbb{E}[\liminf X_n]$. Hence,

$$\mathbb{E}\Big[\liminf_{n\to\infty} X_n\Big] \;\leq\; \liminf_{n\to\infty} \mathbb{E}[X_n].$$

$\square$

Remark. The clever step is defining $Y_n = \min\{Y, X_n\}$. This ensures two things simultaneously:

- $Y_n \leq Y$, so the sequence is uniformly bounded (good for convergence theorems).
- $Y_n \leq X_n$, so we can compare its expectation directly with $\mathbb{E}[X_n]$.

This "best of both worlds" trick is typical in measure-theoretic proofs: construct an auxiliary sequence that inherits the best properties of both sides of the inequality.

Theorem 2.11 (Dominated Convergence Theorem (DCT)). *If $X_n \to X$ a.s. and $|X_n| \leq Y$ for an integrable Y, then*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Proof Idea. Decompose $X_n = X_n^+ - X_n^-$ and apply Fatou's lemma to $Y \pm X_n$. □

Jensen's Inequality

Theorem 2.12 (Jensen's Inequality). *If φ is convex and X integrable, then*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

Proof Sketch. For convex φ, at any x_0 there exists a, b such that $\varphi(x) \geq ax + b$ for all x with equality at x_0. Taking expectations gives

$$\varphi(\mathbb{E}[X]) \leq a\mathbb{E}[X] + b \leq \mathbb{E}[\varphi(X)].$$

□

Example. For $\varphi(x) = x^2$, Jensen's inequality gives $(\mathbb{E}[X])^2 \leq \mathbb{E}[X^2]$, i.e. $(X) \geq 0$.