Lecture Note for MAT8030: Advanced Probability

LI Liying*

September 20, 2024

1 Measure theory preliminaries

In this section we cover some basic facts in measure theory and how they integrate into the modern probability theory, which is essential to this field. Most of the materials are still within the scope of the celebrated work, *Foundations of the theory of probability*, by Kolmogorov in 1933 ([Kol33]).

1.1 Random variables, σ-fields and measures

We start with examples of some random variables (r.v.'s) that the reader should be familiar with from elementary probability. There are two types of r.v.'s encountered in elementary probability: discrete and continuous.

Example 1.1 Examples of discrete r.v.'s.

- **Bernoulli:** $X \sim \mathrm{Ber}(p)$, with $\mathsf{P}(X=1)=p$, $\mathsf{P}(X=0)=1-p$.
- **binomial:** $X \sim \mathrm{Binom}(n,p)$ with $\mathsf{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$, $k=0,1,\dots,n$.
- **geometry:** $X \sim \mathrm{Geo}(p)$, with $\mathsf{P}(X=k)=(1-p)^{k-1}p$, $k=1,2,\dots$.
- **Poisson:** $X \sim \mathrm{Poi}(\lambda)$, with $\mathsf{P}(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$, $k=0,1,\dots$.

Example 1.2 Examples of continuous r.v.'s, described by the density function $\mathsf{P}(X \le a)=\int_{-\infty}^{a} p(x)\,dx$.

- **exponential:** $X \sim \mathrm{Exp}(\lambda)$, with $p(x)=\mathbb{1}_{[0,\infty)}(x)\cdot \lambda e^{-\lambda x}$.
- **uniform:** $X \sim \mathrm{Unif}[a,b]$, with $p(x)=\mathbb{1}_{[a,b]}(x)\cdot \frac{1}{b-a}$.
- **normal/Gaussian:** $X \sim \mathcal{N}(\mu,\sigma^2)$, with $p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}$.

Recall that the distribution/law of a r.v. X is determined by its cumulative distribution function (c.d.f.). In particular, sets of the form $\{X \le a\}$ are *events* of which one can evaluate the probability, denoted by $\mathsf{P}(X \le a)$.

We can say that $\mathsf{P}(\cdot)$ is a function of events, or a *set function*. A measure $\mathsf{P}(\cdot): A \mapsto \mathsf{P}(A) \in [0,\infty)$ is a special set function satisfying the following three properties:

1. **non-negativity**: $\mathsf{P}(A) \ge 0$, $\forall A$.
2. $\mathsf{P}(\varnothing)=0$.

*With contribution from YANG Yuze who typesets some of the note.

3. **countable additivity**: for any *disjoint* $A_1, A_2, \ldots$,

$$\mathsf{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathsf{P}(A_n). \tag{1.1}$$

The last property, countable additivity (a.k.a. σ-additivity) is the most important one. It is only with σ-additivity, not finite additivity, that one can get the hands on various limit theorems for integration/expectation.

Other important properties of measures can be derived from Item 1 to Item 3.

4. **finite additivity** from Items 2 and 3: let $A_{n+1} = A_{n+2} = \cdots = \varnothing$ in (1.1); then

$$\mathsf{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} \mathsf{P}(A_k).$$

5. **monotonicity** from Items 1 and 4: if $A \subset B$, then $A \cap (B \setminus A) = \varnothing$, and hence

$$\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A) \geq \mathsf{P}(A).$$

6. **sub-additivity** from Items 3 and 5: let $\tilde{A}_n = A_n \setminus \big(\bigcup_{k=1}^{n-1} A_k\big) \subset A_n$; then

$$\mathsf{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathsf{P}(\tilde{A}_n) \leq \sum_{n=1}^{\infty} \mathsf{P}(A_n).$$

7. **continuity from above** from Items 2 and 3: if $A_n \downarrow A$ and $\mathsf{P}(A_1) < \infty$, then $\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(A_n)$ ($A = \bigcap_{n=1}^{\infty} A_n$). In fact, since A_1 is the disjoint union of

$$A_1 = A \cup (A_1 \setminus A_2) \cup (A_2 \setminus A_3) \cup \cdots, \tag{1.2}$$

we have

$$\mathsf{P}(A_1) = \mathsf{P}(A) + \mathsf{P}(A \setminus A_n) + \sum_{k=n}^{\infty} \mathsf{P}(A_k \setminus A_{k+1}).$$

All the terms are positive, and the left hand side is finite, so the tail of the infinite sum must converges to 0, and hence

$$\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(A_1) - \mathsf{P}(A \setminus A_n) - \sum_{k=n}^{\infty} \mathsf{P}(A_k \setminus A_{k+1}) = \lim_{n\to\infty} \mathsf{P}(A_1) - \mathsf{P}(A_1 \setminus A_n) = \lim_{n\to\infty} \mathsf{P}(A_n).$$

Note: the decomposition (1.2) *has the following interpretation; as A_n is decreasing, any element $x \in A_1$ either appears in all A_n, and hence in A, or there is a largest n such that $x \in A_n$ but $x \notin A_{n+1}$, and hence $x \in A_n \setminus A_{n+1}$.*

8. **continuity from below** from Items 2, 3, 5 and 7: if $A_n \uparrow A$, then $\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(A_n)$.

Noting that $\mathsf{P}(A_n)$ is increasing, by sub-additivity,

$$\mathsf{P}(A) \leq \mathsf{P}(A_1) + \sum_{n=2}^{\infty} \mathsf{P}(A_n \setminus A_{n-1}) = \lim_{n\to\infty} \mathsf{P}(A_n).$$

If $\mathsf{P}(A) = \infty$, there is nothing else to prove. Otherwise, $\mathsf{P}(A) < \infty$, and $A - A_n \downarrow \varnothing$. Then by continuity from above,

$$0 = \mathsf{P}(\varnothing) = \lim_{n\to\infty} \mathsf{P}(A \setminus A_n) = \lim_{n\to\infty} \mathsf{P}(A) - \mathsf{P}(A_n).$$

We also need to impose some conditions on the domain of the set function $\mathsf{P}(\cdot)$. The domain should behave well under countable union/intersection. This leads to the definition of σ-algebras.

Definition 1.1 *Let Ω be any non-empty set and $\mathcal{F}$ be a collection of subsets of Ω. We say that $\mathcal{F}$ is a σ-algebra (or σ-field), if*

1. $\Omega \in \mathcal{F}$,
2. $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$,
3. *(closure under countable union)* $A_n \in \mathcal{F}$ *implies* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Example 1.3
1. The smallest σ-algebra: $\mathcal{F} = \{\varnothing, \Omega\}$.
2. The largest σ-algebra: $\mathcal{F} = \{$ all subsets of $\Omega\}$.

A set Ω equipped with a σ-algebra $\mathcal{F}$ is called a *measurable space*, written in a pair $(\Omega, \mathcal{F})$.

Proposition 1.1 [1] *Let $\mathcal{F}$ be a σ-field. Then*

- $\varnothing \in \mathcal{F}$,
- $A \subset B$, $A, B \in \mathcal{F}$ *imply* $B \setminus A \in \mathcal{F}$,
- *(closure under countable intersection)* $A_n \in \mathcal{F}$ *implies* $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

Definition 1.2 *A probability space, or probability triple,* $(\Omega, \mathcal{F}, \mathsf{P})$ *is such that* $(\Omega, \mathcal{F})$ *is a measurable space and* $\mathsf{P} : \mathcal{F} \to [0,1]$ *is a measure with* $\mathsf{P}(\Omega) = 1$.

Definition 1.3 *A random variable (r.v.)* $X = X(\omega) : \Omega \to \mathbb{R}$ *is a map from a probability space* $(\Omega, \mathcal{F}, \mathsf{P})$ *to* $\mathbb{R}$, *such that*

$$\{\omega : X(\omega) \leq a\} \in \mathcal{F}, \quad \forall a \in \mathbb{R},$$

or written more compactly, $X^{-1}\big((-\infty, a]\big) \in \mathcal{F}$ *for all* $a \in \mathbb{R}$.

Let us recall some basic facts about the pre-image map φ^{-1} for *any* map $\varphi : U \to V$. It is defined by

$$\varphi^{-1}(W) := \{u \in U : \varphi(u) \in W\}.$$

Proposition 1.2 *The map φ^{-1} commutes with most set operations, in particular:*

- $\varphi^{-1}(W_1 \cap W_2) = \varphi^{-1}(W_1) \cap \varphi^{-1}(W_2)$,
- $\varphi^{-1}(W_1 \cup W_2) = \varphi^{-1}(W_1) \cup \varphi^{-1}(W_2)$,
- $\varphi^{-1}(W^c) = \big(\varphi^{-1}(W)\big)^c$.

Let X be a r.v. on $(\Omega, \mathcal{F}, \mathsf{P})$, and let $\mathcal{B} = \{A \text{ s.t. } X^{-1}(A) \in \mathcal{F}\}$. Definition 1.3 and Proposition 1.2 imply that $\mathcal{B}$ contains all the intervals in $\mathbb{R}$. Moreover, since $\mathcal{F}$ is a σ-algebra,

$$X^{-1}(I_n) \in \mathcal{F} \quad \Longrightarrow \quad X^{-1}\Big(\bigcup_{n=1}^{\infty} I_n\Big) = \bigcup_{n=1}^{\infty} X^{-1}(I_n) \in \mathcal{F}.$$

This implies that $\mathcal{B}$ is also a σ-algebra. As we will see in the next section, $\mathcal{B}$ contains the *Borel σ-algebra*, which is the most important class of σ-algebras in probability theory.

[1] In this note, readers are encouraged to work out their own proofs on propositions without proofs; they are good exercises and will be useful for understanding later materials.

1.2 Construction of σ-algebra and (probability) measures

Simply put, the Borel σ-algebra is the *smallest* σ-algebra containing by open sets. To understand what is "smallest", we start with the following observation.

Lemma 1.3 *1. If $\mathcal{F}_1$ and $\mathcal{F}_2$ are two σ-algebras on Ω, then $\mathcal{F}_1 \cap \mathcal{F}_2$ is also a σ-algebra.*

2. If $\mathcal{F}_\gamma, \gamma \in \Gamma$ are σ-algebras on Ω, where Γ is an arbitrary index set (countable or uncountable), then $\bigcap_{\gamma \in \Gamma} \mathcal{F}_\gamma$ is also a σ-algebra.

Proposition 1.4 *Let $\mathcal{A}$ be a collection of subsets in Ω. Then there exists a smallest σ-algebra containing $\mathcal{A}$, called* the σ-algebra generated by $\mathcal{A}$ *and written $\sigma(\mathcal{A})$, in the sense that if $\mathcal{G} \supset \mathcal{A}$ is a σ-algebra, then $\sigma(\mathcal{A}) \subset \mathcal{G}$.*

Proof: Take $\sigma(\mathcal{A}) = \bigcap_{\mathcal{F}\ \sigma\text{-algebra}:\mathcal{F}\supset\mathcal{A}} \mathcal{F}$. □

Definition 1.4 (Borel σ-algebra) *Let M be a metric space (or any topological space). The* Borel σ-algebra *$\mathcal{B}(M)$ is the σ-algebra generated by all the open sets in M.*

Example 1.4

- $\mathcal{B}(\mathbb{R}) = \sigma\Big((-\infty, a], a \in \mathbb{R}\Big)$.
- $\mathcal{B}(\mathbb{R}^d) = \sigma\Big((-\infty, a_1] \times \cdots \times (-\infty, a_d], a_i \in \mathbb{R}\Big)$.

Remark 1.5 Here, one need to first show that any open sets in $\mathbb{R}^d$ can be obtained from countable union of sets of the form $(-\infty, a_1] \times (-\infty, a_d]$. The construction requires some ideas from point-set topology, but it is elementary, and thus omitted here.

Proposition 1.5 *A map $X(\omega)$ on $(\Omega, \mathcal{F}, \mathsf{P})$ is a r.v. if and only if $X^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{B}(\mathbb{R})$.*

Remark 1.6 In fact, this is usually taken as the definition for r.v.'s.

Now let us take about the distribution of a r.v. X. One can check that $\mu = \mathsf{P} \circ X^{-1}$ defined by

$$\mu(A) = \mathsf{P}(\{\omega : X(\omega) \in A\}), \quad A \in \mathcal{B}(\mathbb{R}),$$

is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We call μ the *distribution/law* of X. Clearly, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ is a probability space. For most of the practical application, say computing expectation, variance, etc, it is enough to understand the distribution of a r.v., not the original probability measure P on some abstract space that can be potentially be very complicate. Another obvious advantage is that the distributions of all r.v.'s are probability measures live on the *same* measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Note that the c.d.f. of a r.v. can be read from its distribution:

$$F_X(a) = \mathsf{P}(X \le a) = \mu\big((-\infty, a]\big), \quad a \in \mathbb{R}.$$

The central topic for this section is to understand how the c.d.f. determines μ. Along the way we will learn how to construct σ-algebras and (probability) measures. Some of the presentation here is taken from [Shi96, Chap. 2.3]. The next theorem is a fundamental and important result.

Theorem 1.6 *Every increasing, right-continuous function $F : \mathbb{R} \to [0, 1]$ with $F(-\infty) = 0$ and $F(\infty) = 1$ uniquely determines a probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.*

We start by introducing some notions on collections of sets.

Definition 1.5 *A collection of sets $\mathcal{S}$ is a* semi-algebra *if first, it is closed under intersection, i.e., $A, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S}$ and second, for every $A \in \mathcal{S}$, A^c is disjoint union of $A_1, A_2, \ldots, A_n$ in $\mathcal{S}$.*

A collection of sets $\mathcal{S}$ is an algebra, *or* field, *if $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$ and $A^c \in \mathcal{S}$.*

These two notions are related by the following proposition.

Proposition 1.7 *Let $\mathcal{S}$ be a semi-algebra. Then*

$$\bar{\mathcal{S}} = \{\textit{finite disjoint unions of sets in } \mathcal{S}\}$$

is an algebra.

Example 1.7 All the d-dimensional half-open, half-closed rectangles forms a semi-algebra:

$$\mathcal{S}_d = \{\varnothing, \quad (a_1, b_1] \times \cdots \times (a_d, b_d], \ -\infty \le a_i < b_i \le \infty\}.$$

Definition 1.6 *A collection of sets $\mathcal{S}$ is a* monotone class (m-class)*, if for every monotone sequence $A_n \in \mathcal{S}$, $A = \lim_{n\to\infty} A_n \in \mathcal{S}$.*

Here, for an increasing sequence $A_n \subset A_{n+1} \subset \cdots$, its limit is defined by $A := \bigcup_{n=1}^{\infty} A_n$, and for an decreasing sequence $A_n \supset A_{n+1} \supset \cdots$, its limit is defined by $A := \bigcap_{n=1}^{\infty} A_n$.

It is easy to see that any *intersection* of m-classes is still an m-class. Therefore, it makes sense to talk about the *smallest* m-classes containing any collection of sets $\mathcal{A}$ (c.f. Proposition 1.4). We denote this smallest m-class by $m(\mathcal{A})$.

The monotone class condition basically bridges the difference between σ-algebras and algebras.

Proposition 1.8 *Let $\mathcal{A}$ be a collection of subsets of Ω. Then $\mathcal{A}$ is a σ-algebra if and only if $\mathcal{A}$ is both an algebra and an m-class.*

Theorem 1.9 *(Monotone Class Theorem) Let $\mathcal{A}$ be an algebra. Then $\sigma(\mathcal{A}) = m(\mathcal{A})$.*

Proof: By Proposition 1.8, $\sigma(\mathcal{A})$ is necessarily an m-class, and by the minimum property we have the inclusion $m(\mathcal{A}) \subset \sigma(\mathcal{A})$.

To show the other direction $\sigma(\mathcal{A}) \subset m(\mathcal{A})$, it suffices to show that $m(\mathcal{A})$ is an algebra, and hence a σ-algebra (using Proposition 1.8 again). To establish that $m(\mathcal{A})$ is an algebra, we will use the *principle of appropriate sets.*

First, $m(\mathcal{A})$ is closed under complement. Let

$$\mathcal{S} = \{A : A, A^c \in m(\mathcal{A})\} \subset m(\mathcal{A}).$$

Our goal is to show that $m(\mathcal{A}) = \mathcal{S}$. Clearly, by definition we have $\mathcal{A} \in \mathcal{S}$. Moreover, $\mathcal{S}$ is an m-class: if $A_n \uparrow A$, $A_n \in \mathcal{S}$, then A_n, A_n^c are both monotone sequence in $m(\mathcal{A})$, and hence their limits $A, A^c \in m(\mathcal{A})$; if $A_n \downarrow A$ it is similar. Therefore, $\mathcal{S}$ must contain the smallest m-class that contains $\mathcal{A}$, which is $m(\mathcal{A})$. This shows $\mathcal{S} = m(\mathcal{A})$, so by the definition of $\mathcal{S}$, $m(\mathcal{A})$ is closed under complement.

Second, $m(\mathcal{A})$ is closed under intersection. Since intersection involves two sets, the proof is slightly more complicated and we will do it in two steps. In the first step, for fixed $A \in \mathcal{A}$, let

$$\mathcal{S}_A = \{B : B \in m(\mathcal{A}), \ A \cap B \in m(\mathcal{A})\} \subset m(\mathcal{A}).$$

Clearly, $\mathcal{A} \subset \mathcal{S}_A$ since A is an algebra and $m(\mathcal{A})$ contains $\mathcal{A}$. Also, $\mathcal{S}_A$ is an m-class as $B_n \downarrow B$ or $B_n \uparrow B$ implies $A \cap B_n \downarrow A \cap B$ or $A \cup B_n \uparrow A \cup B$. Therefore, $m(\mathcal{A}) \subset \mathcal{S}_A$, and we have shown that $A \cap B \in m(\mathcal{A})$ whenever $A \in \mathcal{A}$ and $B \in m(\mathcal{A})$.

In the second step, let

$$\mathcal{S} = \{A \in m(\mathcal{A}) : A \cap B \in m(\mathcal{A}),\ \forall B \in m(\mathcal{A})\}.$$

By the first step, $\mathcal{A} \subset \mathcal{S}$. Again, it is not hard to check that A is an m-class. Therefore $m(\mathcal{A}) = \mathcal{S}$, and this proves that $m(\mathcal{A})$ is closed under intersection.

In conclusion, $m(\mathcal{A})$ is an algebra and hence a σ-algebra, this completes the proof. □

A related concept is the Dynkin system (d-system, λ-class).

Definition 1.7 *Let $\mathcal{D}$ be a collection of subsets of Ω. We say that $\mathcal{D}$ is a Dynkin system if*

1. $\Omega \in \mathcal{D}$,
2. $A, B \in \mathcal{D},\ A \subset B \quad \Rightarrow \quad B \setminus A \in \mathcal{D}$,
3. $A_n \uparrow A,\ A_n \in \mathcal{D} \quad \Rightarrow \quad A \in \mathcal{D}$.

We say that $\mathcal{A}$ is a π-system if it is closed under intersection. One can check that $\mathcal{A}$ is a σ-algebra if and only if it is both a π-system and Dynkin system. Moreover, analogous to Theorem 1.9, the following is true.

Theorem 1.10 *(π-λ Theorem; Dynkin Theorem) If $\mathcal{A}$ is a π-system, then $\sigma(\mathcal{A})$ is the smallest Dynkin system containing $\mathcal{A}$.*

Proof: The proof can be done via the principle of appropriate sets. □

Given a distribution function F as in Theorem 1.6, we can introduce a (probability) measure μ_0 on the algebra

$$\bar{\mathcal{S}} = \{\bigcup_{k=1}^{n} (a_k, b_k],\ \text{disjoint union}\},$$

given by

$$\mu_0(A) = \sum_{k=1}^{n} \big[F(b_k) - F(a_k)\big].$$

It is easy to check that μ_0 is finitely additive. An important step is the following.

Proposition 1.11 *The finitely additive measure μ_0 is σ-additive on $\bar{\mathcal{S}}$, i.e., if $A_n \in \bar{\mathcal{S}}$ are disjoint and $\bigcup_{n=1}^{\infty} A_n \in \bar{\mathcal{S}}$, then*

$$\mu_0(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu_0(A_n).$$

Proof: We will use the fact that σ-additivity is equivalent to continuity at $\varnothing$, i.e., μ_0 is σ-additive if and only if for every $A_n \downarrow \varnothing$, $\lim_{n\to\infty} \mu_0(A_n) = \mu_0(\varnothing) = 0$.

Suppose that there is some $L > 0$ such that $A_n \in [-L, L]$. Let $\varepsilon > 0$. We claim that there exists $B_n \in \bar{\mathcal{S}}$ such that $\overline{B_n} \subset A_n$ and

$$\mu_0(A_n) - \mu_0(B_n) \le \varepsilon \cdot 2^{-n}.$$

The existence of B_n is mostly a consequence of the right continuity of F. In fact, let $A_n = \bigcup_{i=1}^{m} (a_i^{(n)}, b_i^{(n)}]$, and $B_n = \bigcup_{i=1}^{m} (a_i^{(n)} + \delta, b_i^{(n)}]$. Then

$$\mu_0(A_n) - \mu(B_n) = \sum_{i=1}^{m} \big(F(b_i^{(n)} + \delta) - F(b_i^{(n)})\big) \to 0, \quad \delta \downarrow 0.$$

We just need to choose δ small enough so that the sum is less that $\varepsilon \cdot 2^{-n}$.

Since $A_n \downarrow \varnothing$ and $\overline{B_n} \subset A_n$, we have $\overline{B_n} \downarrow \varnothing$. So $C_n = [-L, L] \setminus \overline{B_n}$ forms an open cover of $[-L, L]$. By the Finite Open Cover Theorem, there exists a finite sub-cover, i.e., $\exists n_0$ s.t.

$$[-L, L] \subset \bigcup_{n=1}^{n_0} [-L, L] \setminus \overline{B_n},$$

and hence $\bigcap_{n=1}^{n_0} \overline{B_n} = \varnothing$. Therefore,

$$\mu_0(A_{n_0}) = \mu_0\big(A_{n_0} \setminus \bigcap_{n=1}^{n_0} B_n\big) \leq \mu_0\Big(\bigcup_{n=1}^{n_0} (A_n \setminus B_n)\Big) \leq \sum_{n=1}^{n_0} \mu_0(A_n \setminus B_n) \leq \varepsilon \sum_{n=1}^{\infty} 2^{-n} \leq \varepsilon.$$

Noting that $\mu_0(A_n)$ is decreasing and ε is arbitrary, we have $\lim_{n\to\infty} \mu_0(A_n) = 0$.

For unbounded A_n, since $F(-\infty) = 0$ and $F(\infty) = 1$, for every $\varepsilon > 0$, we can choose L s.t. $\mu_0((-L, L]) \geq 1 - \varepsilon$. Let $\tilde{A}_n = A_n \cap (-L, L]$. Then $\tilde{A}_n \downarrow \varnothing$ and $\tilde{A}_n$ are bounded. Hence, $\lim_{n\to\infty} \mu_0(\tilde{A}_n) = 0$. Therefore,

$$\limsup_{n\to\infty} \mu_0(A_n) \leq \limsup_{n\to\infty} \mu_0(\tilde{A}_n) + \limsup_{n\to\infty} \mu_0\big(A_n \setminus (-L, L]\big) \leq 0 + \varepsilon = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we see that $\lim_{n\to\infty} \mu_0(A_n) = 0$, as desired. □

After establishing Proposition 1.11, we can obtain a unique probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with the help of the next theorem.

Theorem 1.12 (Carathéodory's Extension Theorem) *Let μ_0 be a σ-additive measure on an algebra $\mathcal{A}$. Then μ_0 has a unique extension to $\sigma(\mathcal{A})$.*

Remark 1.8
1. An extension of μ_0 to $\sigma(\mathcal{A})$ is a measure μ on $\sigma(\mathcal{A})$ such that $\mu_0(A) = \mu(A)$ for every $A \in \mathcal{A}$.
2. We will use Theorem 1.12 in the case where μ_0 (and hence the resulting measure) is a *probability* measure. But the theorem also holds when μ_0 is σ-finite, i.e., there exist $A_n \uparrow \Omega$ such that $\mu_0(A_n) < \infty$.

Proof of Uniqueness: Let μ, $\tilde{\mu}$ be two extensions and $\mathcal{S} = \{A : \mu(A) = \tilde{\mu}(A)\}$. We will show (i) $\mathcal{A} \subset \mathcal{S}$; (ii) $\mathcal{A}$ is a monotone class. Then by Theorem 1.9, $\mathcal{S}$ contains $\sigma(\mathcal{A})$ so $\mu = \tilde{\mu}$ on $\sigma(\mathcal{A})$, which proves the uniqueness.

The first statement $\mathcal{A} \subset \mathcal{S}$ follows from definition. To prove the second statement, consider $A_n \uparrow A$ and $A_n \in \mathcal{S}$. Since μ, $\tilde{\mu}$ are measures, from the continuity from below, we have $\mu(A_n) \to \mu(A)$ and $\tilde{\mu}(A_n) \to \tilde{\mu}(A)$, and thus $\mu(A) = \tilde{\mu}(A)$. Similarly, if $A_n \downarrow A$ and $A_n \in \mathcal{S}$, from the continuity from above, we have $\mu(A_n) \to \mu(A)$ and $\tilde{\mu}(A_n) \to \tilde{\mu}(A)$, and thus $\mu(A) = \tilde{\mu}(A)$. □

The existence part makes use of the outer measure. The construction is very much the same as the construction of the Lebesgue measure, so we will only sketch the construction here. Given a measure μ_0 on $\mathcal{A}$, the outer measure is given by

$$\mu_*(A) = \inf\{\sum_{n=1}^{\infty} \mu_0(A_n) : A \subset \bigcup_{n=1}^{\infty} A_n,\ A_n \in \mathcal{A}\}.$$

For the Lebesgue measure, $\mathcal{A}$ consists of nice sets like intervals, rectangles, etc, and we want to use them to generalize the idea of length, area, volume, etc, to Lebesgue measurable sets. A key point is to defined what is "measurable" (w.r.t. the outer measure μ_*). A set A is measurable if it satisfies the Carathéodory's condition:

$$\mu_\star(D) = \mu_\star(D \cap A) + \mu_\star(D \cap A^c), \quad \forall D. \tag{1.3}$$

One can show the following:

1. every set $A \in \mathcal{A}$ satisfies (1.3) and $\mu_*(A) = \mu_0(A)$;

2. the collection of sets that satisfy (1.3), denoted by $\mathcal{F}$, is a σ-algebra and μ_* is a measure on $\mathcal{F}$.

Combining these two statements, one can define

$$\mu := \mu_*\big|_{\sigma(A)},$$

which is the desired extension.

Remark 1.9 Usually $\sigma(\mathcal{A}) \subsetneq \mathcal{F}$. For example, in the Lebesgue measure case $F(x) = x$,

$$\sigma(\mathcal{A}) = \{\text{Borel sets}\}, \quad \mathcal{F} = \{\text{Lebesgue measurable sets}\}.$$

Remark 1.10 If we complete $(\Omega, \sigma(\mathcal{A}), \mu)$, then the result is $(\Omega, \mathcal{F}, \mu_*\big|_{\mathcal{F}})$.
Here, a *complete* probability space $(\Omega, \mathcal{F}, \mathsf{P})$ means that if $B \subset A \in \mathcal{F}$ where $\mathsf{P}(A) = 0$, then $B \in \mathcal{F}$.

1.3 Decomposition of distribution functions

Let $F(x)$ be an increasing, right-continuous function, e.g., the c.d.f. of some r.v. The goal of this section is to decompose it into the jumping (or discontinuous), the absolutely continuous and the singularly continuous parts, written

$$F = F_d + F_{ac} + F_{sc}. \tag{1.4}$$

First we look at the discontinuous part. Since F is right-continuous and increasing, F only has discontinuity points of the first kind. This leads to the following definition.

Definition 1.8 *A point x is a point of jump/discontinuity of F if $F(x) - F(x-) > 0$.*

Proposition 1.13 *The points of jump for an increasing, right-continuous function are countable.*

Proof: On any compact set $[-L, L]$,

$$\{x \in [-L, L] \text{ is a jump}\} = \bigcup_{n=1}^{\infty} \Big\{x \in [-L, L] : F(x) - F(x-) > \frac{1}{n}\Big\}.$$

All sets in the union are finite since each contains at most $n\big(F(L) - F(L-)\big)$ points. The conclusion follows. ☐

Let a_i, $i = 1, 2, \ldots$, be points of jump for the function $F(x)$ and let $b_i = F(a_i) - F(a_i-)$ be the "size of jumps". Define

$$F_d(x) = \sum_{i=1}^{\infty} b_i \mathbb{1}_{[a_i,\infty)}(x).$$

We call F_d the "jumping part". The remaining part $F_c(x) = F(x) - F_d(x)$ is increasing and continuous.
Next we need to classify increasing and continuous functions.

Definition 1.9 (Absolute Continuity) *An increasing, continuous function $F(x)$ is called absolutely continuous if there exist $f \in L^1(\mathbb{R})$ such that $F(b) - F(a) = \int_a^b f(x)\, dx$.*

Remark 1.11 This is generalized Newton–Leibniz formula. By Lebesgue Differentiability Theorem, F' exists almost everywhere and $F' = f$.

The next result is also related, usually proved in real analysis using Vitali covering.

Proposition 1.14 *If F is increasing, then F' exists almost everywhere.*

Note that non-differentiable points in Proposition 1.14 could be points of jumps, but we already took those out. As a result we have the following.

Proposition 1.15 *An increasing and continuous function F can be uniquely decomposed as*

$$F = F_{ac} + F_{sc},$$

where F_{ac} is absolutely continuous and $F_{ac} = \int_{-\infty}^{x} F'(x)\,dx$, and F_{sc} is increasing and continuous but $F'_{sc} \overset{a.e.}{\equiv} 0$.

Remark 1.12 The function F_{sc} appearing in Proposition 1.15 is called "singularly continuous". One may ask if there exists non-trivial singularly continuous function. A famous example is the Cantor's function, or the "Devil's staircase".

Recall that the Cantor set (denoted by $\mathcal{C}$) is constructed by starting with the interval $[0,1] \subset \mathbb{R}$, then dividing it into three intervals of equal length and removing the middle interval, and repeating this process of division and removal. In the end, we obtain

$$\mathcal{C} = [0,1] \setminus \bigcup_{n,k} I_n^{(k)},$$

where $I_n^{(k)}$, $1 \le k \le 2^{n-1}$, $n \ge 1$, are the intervals that are removed in the n-th steps, i.e.,

$$I_1^{(1)} = (\frac{1}{3}, \frac{2}{3}), \quad I_2^{(1)} = (\frac{1}{9}, \frac{2}{9}), \quad I_2^{(2)} = (\frac{7}{9}, \frac{8}{9}), \cdots .$$

Clearly, the Cantor set is a closed set and from a direct calculation of the total length of the removed intervals, the Lebesgue measure of the Cantor set is 0.

The associated Cantor function (denoted by $\varphi(x)$ in this note) is constructed as follows. Set $\varphi(x) = 0$ for $x \le 0$ and $\varphi(x) = 1$ for $x \ge 1$. When $x \in (0,1)$, set $\varphi(x) = \frac{1}{2}$ for $x \in (\frac{1}{3}, \frac{2}{3}) = I_1^{(1)}$, $\varphi(x) = \frac{1}{4}$ for $x \in (\frac{1}{9}, \frac{2}{9}) = I_2^{(1)}$, and $\varphi(x) = \frac{3}{4}$ for $x \in (\frac{7}{9}, \frac{8}{9}) = I_2^{(2)}$, $\dots$. Then define φ on $\mathcal{C}$ by continuity. See also [Dur19, Fig. 1.5].

Proposition 1.16 *There exists a Lebesgue set which is not Borel measurable.*

Proof: We will prove the statement by contradiction. Let $\psi(x) = \frac{1}{2}\big(x + \varphi(x)\big)$ ($\varphi(x)$ is the Cantor function). Then $\psi(x)$ is a continuous, strictly increasing function from $[0,1]$ onto itself. Let $H = \psi^{-1}$ which is also continuous and strictly increasing.

It is easy to check that for any $E \subset [0,1]$,

$$\mathbb{1}_{H(E)}\big(H(x)\big) = \mathbb{1}_E(x).$$

Note that the Lebesgue measure of $\psi(\mathcal{C})$ is $1/2$. Hence, there exists a set $E \subset \psi(\mathcal{C})$ which is NOT Lebesgue measurable. On the other hand, $H(E) = \psi^{-1}(E) \subset \mathcal{C}$ is a subset of Lebesgue measure 0 set, and hence by completeness, it is also Lebesgue measurable. If all Lebesgue sets are Borel measurable, then $\mathbb{1}_{H(E)}$ will be Borel measurable (as the indicator function of a Borel set), and hence $\mathbb{1}_E = \mathbb{1}_{H(E)} \circ H$, being the composition of two Borel measurable functions, is Borel measurable. This gives a contradiction we have chosen E to be non-measurable. $\square$

In the first part of this section, we classify and decompose the distribution functions. In the second part, we will do similar things from the perspective of measures. Let μ be a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition 1.10 *A point x is a point of mass if $\mu(\{x\}) > 0$.*

Let $I = \{x : \mu(\{x\}) > 0\}$ be the set of points of mass. We can define $\mu_d(A) = \sum_{x \in I} \delta_x(A) \cdot \mu(\{x\})$.

$$\delta_x = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

is the *Dirac measure* on x. Then μ_d corresponds to the discrete part of the measure μ. The remaining part $\mu_c = \mu - \mu_d$ will not have point of mass. To further decompose it, we need to introduce the notion of absolute continuity and singularity between measures.

Let P, Q are two probability measures on $(\Omega, \mathcal{F})$ (for the simplest case, one can take $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Definition 1.11 *A measure P is absolute continuous w.r.t. Q, written $P \ll Q$, if $Q(A) = 0$ implies $P(A) = 0$.*

If $P \ll Q$, then by Radon–Nikodym Theorem, there exists a measurable function $f(\omega) \in L^1(Q)$, such that $P(A) = \int_A f(\omega) Q(d\omega)$. We write $f(\omega) = \frac{dP}{dQ}$. The measure Q is called the reference measure. For r.v.'s, the reference measure is the Lebesgue measure.

Definition 1.12 *A r.v. X is continuous if its distribution μ is absolutely continuous with respect to the Lebesgue measure. In this case, the density of X is $\frac{d\mu}{d\mathrm{Leb}}$.*

The last definition is mutual singularity.

Definition 1.13 *Two measures P, Q are mutuality singular, denoted by $P \perp Q$, if there exists A such that $P(A) = 0$ and $Q(A^c) = 0$.*

Example 1.13 Cantor set induce a distribution $\mu_{\mathcal{C}} = \frac{d\varphi}{dt}$, we have $\mu_{\mathcal{C}}\big((\text{Cantor set})^c\big) = 0$ and $\mathrm{Leb}\big(\text{Cantor set}\big) = 0$. Hence $\mu_{\mathcal{C}} \perp \mathrm{Leb}$. In fact, an increasing function F is singularly continuous if $dF \perp \mathrm{Leb}$.

Definition 1.14 *A random variable X is singular if $\mu_x \perp \mathrm{Leb}$.*

How common are singular measures and Cantor-like sets? Surprisingly, they are ubiquitous in probability theory, and usually arise from self-similarities or fractal structures.

Example 1.14 Let $B_t(\omega)$ be the Brownian motion, a random continuous function. For each $a \in \mathbb{R}$, consider its level set

$$\mathcal{Z}_a(\omega) := \{t : B_t(\omega) = a\}.$$

(Note the level set is also random.) For almost every ω and every a, the level $\mathcal{Z}_a(\omega)$ is very similar to a Cantor set, in the sense that it is the complement of the union of nested open intervals, but the interval length can be very random.

Let $B_t^\star = \sup_{0 \le s \le t} B_s$ be the maximal process. Since $t \mapsto B_t$ is continuous, by the definition $B_t^\star$ is increasing and continuous. On the other hand, $\frac{B_t^\star}{dt} \perp \mathrm{Leb}$.

Singular measures can also come from infinite dimensions.

Example 1.15 Let us consider i.i.d. Bernoulli r.v.'s Ber(1/3) and Ber(2/3). More precisely, let $(\Omega, \mathcal{F})$ where

$$\Omega = \{\omega = (\omega_1, \omega_2, \cdots),\ \omega_i \in \{0,1\}\}, \quad \mathcal{F} = \mathcal{P}(\Omega).$$

We can define two probability measures on $(\Omega, \mathcal{F})$:

1. one corresponding to i.i.d. Ber(1/3): $\mathsf{P}_1(\omega_i = 1) = \frac{1}{3}$ and $\mathsf{P}_2(\omega_i = 0) = 2/3$;
2. one corresponding to i.i.d. Ber(2/3): $\mathsf{P}_1(\omega_i = 1) = \frac{2}{3}$ and $\mathsf{P}_2(\omega_i = 0) = 1/3$.

Let

$$A_1 = \{\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \omega_k = \frac{1}{3}\}, \quad A_2 = \{\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n \omega_k = \frac{2}{3}\}.$$

Then by the Strong Law of Large Numbers, we have $\mathsf{P}_1(A_1) = 1$, $\mathsf{P}_2(A_2) = 1$. Clearly, $A_1 \cap A_2 = \emptyset$. It follows that $\mathsf{P}_1(A_2) = 0, \mathsf{P}_2({A_2}^c) = 0$, and hence $\mathsf{P}_1 \perp \mathsf{P}_2$.

1.4 Random variables and measurable maps

Definition 1.15 *A map* $\varphi : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ *is measurable if* $\varphi^{-1}(A) \in \mathcal{F}, \forall A \in \mathcal{S}$.

Definition 1.16 *A r.v.* X *is a measurable map* $(\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. *A random vector* $X = (x_1, \dots, x_d)$ *is a measurable map* $(\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Since the Borel σ-algebra is generated from open sets, we have a simple criterion to check whether a map defines a r.v.

Proposition 1.17 *A map* X *is a random variable if and only if* $X^{-1}(O) \in \mathcal{F}$ *for every open set* O.

Definition 1.17 *A function* f *is a Borel measurable if* f *is measurable map from* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *onto itself.*

Similar to Proposition 1.17, we have the following.

Proposition 1.18 *A function* f *is Borel measurable if and only if* $f^{-1}(O) \subset \mathcal{B}(\mathbb{R})$ *for every open set* O.

To compare with the Lebesgue measurability: f is Lebesgue measurable if and only if $f^{-1}(O)$ is Lebesgue measurable set for every open set O.

Proposition 1.19 *If* f *is Borel measurable and* X *is a random variable, then* $f(X)$ *is a r.v.*

Proof: Let O be a open set. We have $\{\omega : f\big(X(\omega)\big) \in O\} = X^{-1}\big(f^{-1}(O)\big) \in \mathcal{F}$ since $f^{-1}(O) \in \mathcal{B}(\mathbb{R})$. $\square$

Remark 1.16 In this example, if "f is Borel measurable" is replaced by "f is Lebesgue measurable", then the conclusion is false.

Proposition 1.20 *If* $f : \mathbb{R} \to \mathbb{R}^d$ *is a Borel map and* $X = (x_1, \cdots, x_d)$ *is a random vector, then* $f(X) = f(x_1, \cdots, x_d)$ *is a random variable.*

Example 1.17 $X_1 + X_2$, $\min\{X_1, X_2\}$ are also random variables.

Proposition 1.20 can help us generate many random variables. Next we need to understand the limits of r.v.'s.

Proposition 1.21 *Let* $X_n, n = 1, 2, \dots$ *be random variables, then*

$$\sup_{n\geq 1} X_n, \quad \inf_{n\geq 1} X_n, \quad \limsup_{n\to\infty} X_n, \quad \liminf_{n\to\infty} X_n$$

are random variables.

Proof:

(i) Let $Y_1(\omega) = \sup_n X_n(\omega)$. We need to show that $Y_1^{-1}((-\infty, a]) \in \mathcal{F}$. For every $a \in \mathbb{R}$,

$$Y_1^{-1}((-\infty, a]) = \{\omega : \sup_n X_n(\omega) \leq a\} = \bigcap_{n=1}^{\infty} \{\omega : X_n(\omega) \leq a\} \in \mathcal{F}.$$

Therefore, $Y_1(\omega) = \sup_n X_n(\omega)$ is a r.v.

(ii) Let $Y_2(\omega) = \inf_n X_n(\omega)$. We need to show that $Y_2^{-1}([a,\infty)) \in \mathcal{F}$. For every $a \in \mathbb{R}$,

$$Y_2^{-1}([a,-\infty)) = \{\omega : \inf_n X_n(\omega) \geq a\} = \bigcap_{n=1}^{\infty} \{\omega : X_n(\omega) \geq a\} \in \mathcal{F}.$$

Therefore, $Y_2(\omega) = \inf_n X_n(\omega)$ is a r.v.

(iii) Since for every ω,

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

we have for every $a \in \mathbb{R}$,

$$\begin{aligned}
\{\omega : \limsup_{n\to\infty} X_n(\omega) > a\} &= \{\omega : \inf_{n\geq 1} \sup_{m\geq n} X_m(\omega) > a\} \\
&= \bigcap_{n=1}^{\infty} \{\omega : \sup_{m\geq n} X_m(\omega) > a\} \\
&= \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{\omega : X_m(\omega) > a\}
\end{aligned}$$

(iv) Since for every ω,

$$\liminf_{n\to\infty} X_n(\omega) = \sup_{n\geq 1} \inf_{m\geq n} X_m(\omega),$$

we have for every $a \in \mathbb{R}$,

$$\begin{aligned}
\{\omega : \liminf_{n\to\infty} X_n(\omega) < a\} &= \{\omega : \sup_{n\geq 1} \inf_{m\geq n} X_m(\omega) < a\} \\
&= \bigcap_{n=1}^{\infty} \{\omega : \inf_{m\geq n} X_m(\omega) < a\} \\
&= \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{\omega : X_m(\omega) < a\}.
\end{aligned}$$

$\square$

Corollary 1.22 *Let X_n, $n = 1, 2, \ldots$, be r.v.'s. The set $\Omega_0 = \{\omega : \lim_{n\to\infty} X_n(\omega)$ exists$\}$ belongs to $\mathcal{F}$.*

Proof: Note that

$$\Omega_0 = \{\omega : \lim_{n\to\infty} X_n(\omega)\} = \{\omega : \limsup_{n\to\infty} X_n(\omega) - \liminf_{n\to\infty} X_n(\omega) = 0\}.$$

By Proposition 1.21, $Y_1 = \limsup_{n\to\infty} X_n(\omega)$ and $Y_2 = \liminf_{n\to\infty} X_n(\omega)$ are r.v.'s, and hence $Y_1 - Y_2$ is a r.v.. Therefore, $\Omega_0 = \{Y_1 - Y_2 = 0\} \in \mathcal{F}$. $\square$

1.5 Integration and expectation

In this section we brief present the theory of integration of measurable functions, or in the context of probability theory, the mathematical expectation. The main difference is that in probability theory, we mostly deal with a finite measure (the total mass of a probability measure is 1). Let X be a random variable on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$. We will denote its expectation X by $\mathsf{E}(X)$. Using a more measure theory oriented notation, sometimes we also write

$$\mathsf{E}X = \int X(\omega)\, \mathsf{P}(d\omega). \tag{1.5}$$

The definition of (1.5) is through approximation via simple random variables (a.k.a. simple functions in measure theory).

We say that a r.v. $X(\omega)$ is *simple*, if there exists finitely many $A_1, \ldots, A_n \in \mathcal{F}$ and $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$X(\omega) = \sum_{k=1}^{n} c_k \mathbb{1}_{A_k}(\omega). \tag{1.6}$$

In the case of (1.6), unquestionably we should define

$$\mathsf{E}(X) = \sum_{k=1}^{n} c_k \mathsf{P}(A_k). \tag{1.7}$$

It is routine to verify common integral properties for expectation of simple r.v.'s, e.g., linearity, monotonicity, order preserving, etc, so we omit it in this note.

Using approximation, we can define expectation of general r.v.'s. For a positive r.v. $X(\omega) \geq 0$, we define

$$\mathsf{E}X = \int_\Omega X(\omega)P(d\omega) := \sup\{\int Y(\omega)P(d\omega) : \text{Y is simple } 0 \leq Y(\omega) \leq X(\omega)\} \in [0, \infty]. \tag{1.8}$$

For general X, $X(\omega) = X_+(\omega) - X_-(\omega)$ where $X_+(\omega) = X(\omega)\mathbb{1}_{\{X>0\}}$ and $X_-(\omega) = -X(\omega)\mathbb{1}_{\{X\leq 0\}}$ are the positive and negative parts of X. If $\mathsf{E}(X_+) < \infty$ or $\mathsf{E}(X_-) < \infty$, then $\mathsf{E}(X) = \mathsf{E}(X_+) - \mathsf{E}(X_-)$. Otherwise, $\mathsf{E}X$ is undefined since X we cannot define $\infty - \infty$.

Next, we want to discuss conditions that justifies exchanging order of taking limit and expectation, i.e.,

$$\mathsf{E} \lim_{n\to\infty} X_n = \lim_{n\to\infty} \mathsf{E}X_n. \tag{1.9}$$

Lemma 1.23 *Suppose that $X_n \geq 0$ are simple and $X_n \uparrow X$. Then* (1.9) *holds.*

Remark 1.18 If "$X_n \uparrow X$" is replaced by "$X_n \leq X$ and $X_n \to X$", we can still an get increasing sequence by considering $Y_n = \max_{1\leq k\leq n} X_k$. It is easy to see that Y_n are also simple and $Y_n \uparrow X$.

Proof: Form the definition (1.8), we have $\mathsf{E}(X) \geq \mathsf{E}(X_n)$.

If $\mathsf{E}X < \infty$, we need to show $\mathsf{E}(X) \leq \lim_{n\to\infty} \mathsf{E}(X_n)$. For every $\varepsilon > 0$, by the definition of supremum, there exists a simple function Y_ε such that $Y_\varepsilon \leq X$ and $\mathsf{E}(Y_\varepsilon) \geq \mathsf{E}(X) - \varepsilon$. For every $\delta > 0$, let $A_n = \{\omega : X(\omega) \leq Y_\varepsilon(\omega) - \delta\}$. Since $X_n(\omega) \uparrow X(\omega) \geq Y_\varepsilon(\omega)$, we have $A_n \uparrow \Omega$ and hence $A_n^c \downarrow \emptyset$. We have

$$\begin{aligned}
\mathsf{E}X_n &= \mathsf{E}X_n\mathbb{1}_{A_n} + \mathsf{E}X_n\mathbb{1}_{A_n{}^c} \geq \mathsf{E}(Y_\varepsilon - \delta)\mathbb{1}_{A_n} \\
&= \mathsf{E}Y_\varepsilon\mathbb{1}_{A_n} - \delta P(A_n) \\
&= \mathsf{E}Y_\varepsilon - \mathsf{E}Y_\varepsilon\mathbb{1}_{A_n{}^c} - \delta P(A_n) \\
&\geq \mathsf{E}X - \varepsilon - \sup_\omega Y_\varepsilon(\omega) \cdot P(A_n{}^c) - \delta
\end{aligned}$$

Since Y_ε is simple, it is always bounded and $\sup_\omega Y_\varepsilon(\omega) < \infty$. Letting $n \to \infty$, we have $\mathsf{E}X_n \geq \mathsf{E}X - \varepsilon - \delta$. Since $\varepsilon, \delta > 0$ are arbitrary, we have $\lim_{n\to\infty} \mathsf{E}X_n \geq \mathsf{E}X$.

If $\mathsf{E}X = \infty$, we need to show $\lim_{n\to\infty} \mathsf{E}(X_n) = \infty$. Again, by (1.8), $\forall M > 0$, there exists a simple function Y_M satisfies $Y_M \leq X$ and $\mathsf{E}(Y_M) \geq M$. For every $\xi > 0$, let $B_n = \{\omega : X(\omega) \geq Y_M(\omega) - \xi\}$. Since $X_n(\omega) \uparrow X(\omega) \geq Y_M(\omega)$, we have $B_n \uparrow \Omega$ and hence $B_n^c \downarrow \emptyset$. Therefore,

$$\begin{aligned}
\mathsf{E}X_n &= \mathsf{E}X_n \mathbb{1}_{B_n} + \mathsf{E}X_n \mathbb{1}_{{B_n}^c} \geq \mathsf{E}(Y_M - \xi)\mathbb{1}_{B_n} \\
&= \mathsf{E}Y_M \mathbb{1}_{B_n} - \xi P(B_n) \\
&= \mathsf{E}Y_M - \mathsf{E}Y_M \mathbb{1}_{{B_n}^c} - \xi P(B_n) \\
&\geq M - \sup_\omega Y_M(\omega) P({B_n}^c) - \xi
\end{aligned}$$

Letting $n \to \infty$, we have $\mathsf{E}X_n \geq M - \xi$. Since $M, \xi > 0$ are arbitrary, we have $\lim_{n\to\infty} \mathsf{E}X_n = \infty$. $\square$

Note that for any non-negative r.v. X, one can find simple r.v.'s $X_n \uparrow X$ so that Lemma 1.23 applies. A simple construction is

$$X_n(\omega) = \frac{\lfloor 2^n X(\omega) \rfloor}{2^n} \wedge n = \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} \mathbb{1}_{\{X(\omega) \in [\frac{k}{2^n}, \frac{k+1}{2^n})\}} + n\mathbb{1}_{\{X(\omega) \geq n\}},$$

where $a \wedge b := \min(a, b)$. To see that $X_n \to X$, notice that if $X(\omega) \leq n$, then $|X(\omega) - X_n(\omega)| \leq \frac{1}{2^n}$.

Theorem 1.24 (Monotone Converge Theorem) *If $X_n \geq 0$ and $X_n \uparrow X$, then* (1.9) *holds.*

Proof: Let $Y_n^{(m)}$, $n, m = 1, 2, \ldots$, be simple functions such that $X_n = \lim_{n\to\infty} Y_n^{(m)}$, and let $Z^{(m)} = \max_{1\leq n\leq m} Y_n^{(m)}$, i.e. $Z^{(1)} = Y_1^{(1)}$, $Z^{(2)} = \max\{Y_1^{(2)}, Y_2^{(2)}\}$, $Z^{(3)} = \max\{Y_1^{(3)}, Y_2^{(3)}, Y_3^{(3)}\}$, $\cdots$. Then $Z^{(m)}$ are also simple and $Z^{(m)} \leq X$,$m = 1, 2, \ldots$. Moreover, we have

$$Y_n^{(m)} \leq Z^{(m)} \leq X_m, \quad \forall m \geq n. \tag{1.10}$$

Taking $m \to \infty$ in (1.10), we see that $X_n \leq \lim_{m\to\infty} Z^{(m)} \leq X$ for every n. Letting $n \to \infty$ we see that $\lim_{m\to\infty} Z^{(m)} = X$. In summary, $Z^{(m)} \uparrow X$ and $Z^{(m)}$ are simple.

Using Lemma 1.23 we have

$$\mathsf{E}X = \lim_{m\to\infty} \mathsf{E}Z^{(m)} \leq \liminf_{m\to\infty} \mathsf{E}X_m.$$

Since $X_m \uparrow X$ we have $\mathsf{E}X_m \leq \mathsf{E}X$ and hence $\limsup_{m\to\infty} \mathsf{E}X_m = X$. Therefore, $\mathsf{E}X = \lim_{m\to\infty} \mathsf{E}X_m$ and the proof is complete.

$\square$

References

[Dur19] Richard Durrett. *Probability: Theory and Examples.* Number 49 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, fifth edition edition, 2019.

[Kol33] A.N. Kolmogorov. *Foundations of the Theory of Probability (English Translation).* 1933.

[Shi96] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics.* Springer New York, 1996.