MATH5635 Stochastic Calculus for Finance I Autumn 2025

Chapter 2: Integrations and Convergence

Lecturer: Kenneth Ng

Preview

This chapter introduces the Lebesgue integral, which is an integral with respect to a measure. In particular, expected values of random variables are defined in the Lebesgue sense. We first highlight the 3-step constructions of Lebesgue integrals for integrable functions. We then discuss different modes of convergence of random variables – almost sure convergence, L^p convergence, and convergence in probability. We then introduce the fundamental convergence theorems of integrals, monotone convergence and dominated convergence.

Key topics in this chapter:

1. Expected values as Lebesgue integrals;
2. Constructions of Lebesgue integrals;
3. Convergence of random variables;
4. Convergence of integrals.

1 Lebesgue and Riemann Integrals

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and μ be a *σ-finite* measure[1]. Let $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a measurable function. We want to construct the ***Lebesgue integral*** of f with respect to μ, denoted by

$$\mu(f) = \int_\Omega f(x)\, d\mu(x).$$

In particular, if $\mu = \mathbb{P}$ is a probability measure, and $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable, the Lebesgue integral of X with respect to $\mathbb{P}$ is called the ***expected value of*** X, denoted by

$$\mathbb{E}[X] := \int_\Omega X(\omega)\, d\mathbb{P}(\omega).$$

[1]Recall that a measure is σ-finite if there exists a collection $\{A_n\}_{n=1}^\infty \subseteq \mathcal{F}$ with $\cup_{n=1}^\infty A_n = \Omega$, and $\mu(A_n) < \infty$ for each $n \in \mathbb{N}$. In particular, a finite measure (e.g., probability measure) is also σ-finite.

When $\Omega = [a, b]$ for some $a < b$, $\mathcal{F} = \mathcal{B}([a, b])$, and $\mu = \lambda$, the Lebesgue measure, we also write the Lebesgue integral of a function f with respect to λ as

$$\int_a^b f(x)\, d\lambda(x) = \int_a^b f(x)\, dx.$$

In elementary calculus courses, the expression $\int_a^b f(x)\, dx$ often refers to the *Riemann integral*, which is defined as follows:

Definition 1.1 (Riemann Integral) Let $f : [a, b] \to \mathbb{R}$ be a bounded function. A partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ satisfies $a = x_0 < x_1 < \cdots < x_n = b$. Define

$$\Delta x_i = x_i - x_{i-1}, \quad m_i = \inf_{x \in [x_{i-1}, x_i]} f(x), \quad M_i = \sup_{x \in [x_{i-1}, x_i]} f(x).$$

Then the lower and upper Riemann sums are given by

$$L(f, P) := \sum_{i=1}^n m_i \Delta x_i, \quad U(f, P) := \sum_{i=1}^n M_i \Delta x_i.$$

We say f is **Riemann integrable** on $[a, b]$ if

$$-\infty < \sup_P L(f, P) = \inf_P U(f, P) < \infty,$$

and in that case, we define

$$\int_a^b f(x)\, dx := \sup_P L(f, P) = \inf_P U(f, P).$$

The Lebesgue integral is a more general integration theory that applies to a broader class of functions and measures. In particular, every Riemann integrable function is Lebesgue integrable, and the two integrals agree when both are defined. However, the converse is not true. We present a famous example of function that is not Riemann integrable on $[0, 1]$, but we will show in the sequel that it is Lebesgue integrable.

Example 1.1 Define $f : [0, 1] \to \mathbb{R}$ by $f(x) = \mathbb{1}_{\mathbb{Q}}(x)$. For any partition $P = \{x_0, x_1, \ldots, x_n\}$, using the density of rational numbers, we must have $m_i = 0$ and $M_i = 1$ for any $i = 1, \ldots, n$. Hence,

$$0 = \sup_P L(f, P) < \inf_P U(f, P) = 1.$$

Hence, f is not Riemann integrable. This is because f has infinitely (despite countably) many discontinuities.

2 Construction of Lebesgue Integrals

We construct a Lebesgue integral of f using the following 3-step procedure:

1. f is a simple function;
2. f is a non-negative function;
3. f is a general measurable function.

The idea behind this procedure is to begin by defining the Lebesgue integral for the simplest class of functions. We then approximate more general functions using these simpler ones, and define the integral of the general function as the limit of the integrals of its approximations. A similar strategy will be used later in this course when we construct stochastic integrals.

2.1 Simple Functions

Recall that a measurable function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called *simple* if it can be written as

$$f(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(x), \tag{1}$$

where $a_1, \dots, a_n \in \mathbb{R}$, $A_1, \dots, A_n \in \mathcal{F}$, and $A_i \cap A_j = \varnothing$.

The Lebesgue integral for a simple function is defined as follows.

Definition 2.1 Let $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a simple function taking the form (1). Then, the Lebesgue integral of f with respect to μ is defined as

$$\int_\Omega f(x) d\mu(x) := \sum_{i=1}^{n} a_i \mu(A_i).$$

Below are some properties of Lebesgue integrals for simple functions. We use the shorthand notation $\int f \, d\mu$ to denote $\int_\Omega f(x) d\mu(x)$.

Proposition 2.1 Let f and g be two simple functions. The following properties hold:

1. for any $a, b \in \mathbb{R}$, $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$;
2. if $f \geq g$ μ-a.e., then $\int f d\mu \geq \int g d\mu$;
3. if $f = g$ μ-a.e., then $\int f d\mu = \int g d\mu$;
4. $|\int f d\mu| \leq \int |f| d\mu$.

Proof. 1. By definition, it is clear that $\int af \, d\mu = a \int f \, d\mu$ and $\int bg \, d\mu = b \int g \, d\mu$. Hence, it suffices to show that $\int (f+g)d\mu = \int f d\mu + \int g d\mu$. Let

$$f(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(x) \text{ and } g(x) = \sum_{j=1}^{m} b_j \mathbb{1}_{B_j}(x).$$

Let $a_o = b_0 = 0$, $A_0 := \cup_{j=1}^{m} B_j \setminus \cup_{i=1}^{n} A_i$ and $B_0 := \cup_{i=1}^{n} A_i \setminus \cup_{j=1}^{m} B_j$. Then,

$$f(x) + g(x) = \sum_{i=0}^{n} \sum_{j=0}^{m} (a_i + b_j) \mathbb{1}_{A_i \cap B_j}(x).$$

Note that $A_i \cap B_j$ are pairwise disjoint, $\cup_{i=0}^{n} A_i = \cup_{j=0}^{m} B_j = (\cup_{i=1}^{n} A_i) \cup (\cup_{j=1}^{m} B_j)$, $\cup_{j=0}^{m} (A_i \cap B_j) = A_i$, and $\cup_{i=0}^{m} (A_i \cap B_j) = B_j$. Hence,

$$\begin{aligned}
\int_\Omega (f(x) + g(x)) d\mu(x) &= \sum_{i=0}^{n} \sum_{j=0}^{m} (a_i + b_j) \mu(A_i \cap B_j) \\
&= \sum_{i=0}^{n} a_i \sum_{j=0}^{m} \mu(A_i \cap B_j) + \sum_{j=0}^{m} b_j \sum_{i=0}^{n} \mu(A_i \cap B_j) \\
&= \sum_{i=0}^{n} a_i \mu(A_i) + \sum_{j=0}^{m} b_j \mu(B_j) \\
&= \sum_{i=1}^{n} a_i \mu(A_i) + \sum_{j=1}^{m} b_j \mu(B_j) \\
&= \int_\Omega f(x) d\mu(x) + \int_\Omega g(x) d\mu(x).
\end{aligned}$$

2. By the definition of Lebesgue integral, it is clear that $\int \varphi d\mu \geq 0$ if $\varphi \geq 0$ μ-a.e. Using this and the first property, $\int f\, d\mu = \int (f-g) d\mu + \int g\, d\mu \geq \int g\, d\mu$, since $f - g = \varphi \geq 0$ μ-a.e.

3. $f = g$ μ-a.e. means $f \geq g$ μ-a.e. and $g \geq f$ μ-a.e. Using the second property, we have $\int f\, d\mu \geq \int g\, d\mu$ and $\int f\, d\mu \leq \int g\, d\mu$, which implies $\int f\, d\mu = \int g\, d\mu$.

4. Note that $-|f| \leq f \leq |f|$ μ-a.e. Using the second property, we have $-\int |f| d\mu \leq \int f\, d\mu \leq \int |f| d\mu$, which implies $|\int f\, d\mu| \leq \int |f| d\mu$.

□

2.2 Non-negative Functions

Next, we construct the Lebesgue integral for non-negative measurable functions f, i.e., $f \geq 0$ μ-a.e.

We first show that any non-negative function f can be approximated by a sequence of simple functions f_n such that $f_n \le f_{n+1}$ and $f_n \to f$ μ-a.e. (see Section 4 for the definition of almost-everywhere convergence). To see this, for any $n \in \mathbb{N}$ and $k \le n2^n$, we define the (measurable) set

$$A_{n,k} := \left\{ \frac{k-1}{2^n} \le f(x) < \frac{k}{2^n} \right\}, \; A_{n,n2^n+1} := \{f(x) \ge n\}.$$

Then, we define f_n by

$$f_n(x) := \sum_{k=1}^{n2^n} \frac{k-1}{2^n} \mathbb{1}_{A_{n,k}}(x) + n\mathbb{1}_{A_{n,n2^n+1}}(x). \tag{2}$$

By constructions, each f_n is a positive simple function and $f_n \le f_{n+1}$. This approximation scheme is also called a ***dyadic approximation***.

Since each f_n is a simple function, we can define $\int f_n \, d\mu$. By Property 2 of Proposition 2.1, we have

$$0 \le \int_\Omega f_n(x) \, d\mu(x) \le \int_\Omega f_{n+1}(x) \, d\mu(x).$$

By the *monotone convergence theorem* of non-negative sequence, we have

$$\lim_{n\to\infty} \int_\Omega f_n(x) \, d\mu(x) = \sup_{n\in\mathbb{N}} \int_\Omega f_n(x) \, d\mu(x) \in [0, \infty].$$

We define the Lebesgue integral of f as this limit:

Definition 2.2 Let f be a non-negative measurable function, and let $\{f_n\}_{n=1}^\infty$ be a sequence of dyadic simple functions approximating f from below, as defined in (2). Then, we define the Lebesgue integral of f by

$$\int_\Omega f(x) \, d\mu(x) := \lim_{n\to\infty} \int_\Omega f_n(x) \, d\mu(x).$$

The properties stated in Proposition 2.1 also hold for the Lebesgue integrals of general non-negative measurable functions f and g. This follows from approximating f and g by sequences of simple functions $\{f_n\}$ and $\{g_n\}$, respectively, each satisfying these properties. By passing to the limit (using monotone convergence), the functions f and g inherit the same integral properties.

Proposition 2.2 The properties about the integrals $\int f \, d\mu$ and $\int g \, d\mu$ in Proposition 2.1 still hold if f, g are non-negative measurable functions.

2.3 General Functions

Finally, we construct the Lebesgue integral for a general measurable function which is neither positive nor simple.

To this end, define

$$f^+(x) := \max\{f(x), 0\} \text{ and } f^- := \max\{-f(x), 0\},$$

which are the positive and negative part of f, respectively. Note that $f^+, f^- \geq 0$, and $f = f^+ - f^-$. From the last subsection, we have constructed Lebesgue integrals for positive functions, so that $\int f^+ \, d\mu$ and $\int f^- \, d\mu$ are well-defined. These observations motivate us to define $\int f \, d\mu$ as follows

Definition 2.3 The Lebesgue integral of a measurable function f is defined as

$$\int_\Omega f(x) \, d\mu(x) := \int_\Omega f^+(x) \, d\mu(x) - \int_\Omega f^-(x) \, d\mu(x).$$

Again, the Lebesgue integrals for general measurable functions also satisfy the properties in Proposition 2.1.

Proposition 2.3 The properties about the integrals $\int f \, d\mu$ and $\int g \, d\mu$ in Proposition 2.1 still hold if f, g are general measurable functions.

2.4 Riemann Integrals vs Lebesgue Integrals

We have seen from Example 1.1 that, even the space $\Omega = [0, 1]$ is compact, a bounded measurable function may not be Riemann integrable. In contrast, a bounded measurable function is always Lebesgue integrable. We first look at the Lebesgue integral of $f = \mathbb{1}_{\mathbb{Q}}$ in Example 1.1.

Example 2.1 Continuing from Example 1.1 with $(\Omega, \mathcal{F}) = ([0, 1], \mathcal{B}([0, 1]))$, and $f(x) := \mathbb{1}_{\mathbb{Q}}(x)$. Calculate the Lebesgue integral $\int_0^1 f(x)dx$.

Solution. Note that f is a simple function. By definition,

$$\int_0^1 f(x) \, dx = \int_0^1 \mathbb{1}_{\mathbb{Q}}(x) \, dx = \lambda(\mathbb{Q} \cap [0, 1]),$$

where $\lambda(\mathbb{Q})$ is the Lebesgue measure of the rational number set in $[0, 1]$. Since $\mathbb{Q}$ is countable, using the countable additivity of measures, we have

$$\lambda(\mathbb{Q} \cap [0, 1]) = \sum_{q \in \mathbb{Q} \cap [0,1]} \lambda(\{q\}) = \sum_{q \in \mathbb{Q} \cap [0,1]} 0 = 0,$$

since the Lebesgue measure of a point is zero; recall Example 2.1 in Chapter 1. Hence, $\int_0^1 f(x)\,dx = 0$. Indeed, $f = 0$ λ-a.e. $\square$

The following result relates Riemann and Lebesgue integrability. In particular, if a function $f : [a, b] \to \mathbb{R}$ is Riemann integrable, it must be Lebesgue integrable, and the two notions agree. Hence, we can use the usual integration rule to compute a Lebesgue integral.

Theorem 2.4 Let $f : [a, b] \to \mathbb{R}$ be a bounded function.

1. The Riemann integral $\int_a^b f(x)\,dx$ exists if and only if the set of discontinuities of f has Lebesgue measure zero, i.e., f is continuous almost everywhere.
2. If the Riemann integral $\int_a^b f(x)\,dx$ exists, then f is Borel measurable, the Lebesgue integral $\int_{[a,b]} f(x)\,dx$ is defined, and both integrals are equal.

3 Expected Values and Properties

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable. Recall that the Lebesgue integral of X with respect to $\mathbb{P}$ is simply the expected value of X, defined by

$$\mathbb{E}[X] := \int_\Omega X(\omega) d\mathbb{P}(\omega).$$

In this section, we focus on properties of $\mathbb{E}[X]$. Most of the properties introduced in this section also holds for Lebesgue integrals with respect to a σ-finite measure.

3.1 L^p Space

Definition 3.1

1. A random variable X is said to be integrable if $\mathbb{E}[|X|] < \infty$. We write $L^1(\Omega, \mathcal{F}, \mathbb{P})$ (or simply L^1) to denote the set of all integrable random variables.
2. For $1 \leq p < \infty$, a random variable is said to be L^p-integrable if $|X|^p \in L^1$.
3. For $p = \infty$, we say that $X \in L^\infty$ if there exists $M > 0$ such that $|X| \leq M$ a.s.

Remark 3.1. For $p \in [1, \infty)$,

1. L^p is a vector space. In particular, if $X, Y \in L^p$, then $X + Y \in L^p$.
2. L^p is a *normed space* equipped with the p-norm, defined by

$$\|X\|_p := (\mathbb{E}[|X|^p])^{\frac{1}{p}}.$$

If $p = \infty$, L^∞ is a normed space equipped with the ∞-norm

$$\|X\|_\infty := \operatorname*{ess\,sup}_{\omega \in \Omega} |X(\omega)| := \inf\{M \geq 0 : |X| \leq M \text{ a.s.}\}.$$

3.2 Some Important Inequalities

We introduce three fundamental inequalities regarding expected values.

Theorem 3.2 (Chebyshev's inequality) Let $X \in L^p$. Then, for any $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|^p \mathbb{1}_{\{|X| \geq a\}}]}{a^p} \leq \frac{\mathbb{E}[|X|^p]}{a^p}.$$

Proof. Since $|X|^p \geq |X|^p \mathbb{1}_{\{|X| \geq a\}} \geq a^p \mathbb{1}_{\{|X| \geq a\}}$. By taking expectations on all sides,

$$a^p \mathbb{P}(|X| \geq a) \leq \mathbb{E}[|X|^p \mathbb{1}_{\{|X| \geq a\}}] \leq \mathbb{E}[|X|^p].$$

The result follows by dividing all sides by a^p. □

For the next two inequalities, the proofs are provided in the Appendix A.

Theorem 3.3 (Hölder's inequality) Let $p, q \in [1, \infty]$ such that $1/p + 1/q = 1$. Let $X \in L^p$ and $Y \in L^q$. Then,

$$\mathbb{E}[|XY|] \leq \|X\|_p \|Y\|_q.$$

In particular, Hölder's inequality is reduced to the ***Cauchy–Schwarz*** inequality if $p = q = 2$.

Theorem 3.4 (Jensen's inequality) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function, i.e., for any $\lambda \in [0, 1]$ and $x, y \in \mathbb{R}$,

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)\varphi(y).$$

Then, for any random variable X such that $X \in L^1$ and $\varphi(X) \in L^1$,

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

Remark 3.5.

1. Chebyshev's inequality holds for a general σ-finite measure.
2. Hölder's inequality holds for a general σ-finite measure for $p, q \in (1, \infty)$.
3. If μ is a finite measure, we have the following generalization of Jensen's inequality:

$$\varphi\left(\frac{1}{\mu(\Omega)} \int_\Omega f(x)\, d\mu(x)\right) \leq \frac{1}{\mu(\Omega)} \int_\Omega \varphi \circ f(x)\, d\mu(x).$$

   When $\mu(\Omega) = 1$, i.e., μ is a probability measure, we obtain the version as in Theorem 3.4.

Using Jensen's inequality, we can show the nested structure of L^p spaces under finite measures.

Theorem 3.6 Let $p > q \geq 1$. Then, $L^p \subseteq L^q$, i.e., any L^p-integrable random variables must be L^q-integrable if $q \leq p$.

Proof. If $p = \infty$, then there exists $M > 0$ such that $|X| \leq M$ a.s. This implies that $\mathbb{E}[|X^q|] \leq \mathbb{E}[M^q] = M^q < \infty$, whence $X \in L^q$.

If $p < \infty$, let $\varphi(x) := x^{p/q}$, which is a convex function since $p > q$. Let $Y = |X|^q$, so that $|X|^p = Y^{p/q} = \varphi(Y)$. Using the fact that $X \in L^p$,

$$\infty > \mathbb{E}[|X|^p] = \mathbb{E}[\varphi(Y)] \geq \varphi(\mathbb{E}[Y]) = \varphi(\mathbb{E}[|X|^q]) = (\mathbb{E}[|X|^q])^{\frac{p}{q}},$$

which implies $\mathbb{E}[|X|^q] < \infty$.

$\square$

4 Convergence of Random Variables

In Section 2.2, we introduced the dyadic approximation (2) which "converges" to f as $n \to \infty$, although the notion of convergence is not obvious in the case of measurable functions or random variables. In this section, we introduce three modes of convergence and discuss their relationship.

The first type of convergence is called *almost-sure convergence*, which is a type of pointwise convergence except in measure zero sets.

Definition 4.1 (Almost-sure convergence) The sequence of random variables $\{X_n\}_{n=1}^\infty$ is said to ***converge almost-surely*** to X if

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

for almost all $\omega \in \Omega$, i.e., $\mathbb{P}(\lim_{n\to\infty} X_n \neq X) = 0$. In this case, we write $X_n \to X$ a.s.

The second type of convergence is defined in terms of the integrals of the deviation $|X_n - X|^p$, $p \geq 1$.

Definition 4.2 (L^p convergence) Given that $\{X_n\}_{n=1}^\infty \subseteq L^p$ and $X \in L^p$, where $1 \leq p < \infty$. The sequence $\{X_n\}_{n=1}^\infty$ is said to be ***convergent in*** L^p to X if

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0.$$

In this case, we write $X_n \xrightarrow{L^p} X$.

Remark 4.1. Let $p \geq q \geq 1$, $\{X_n\}_{n=1}^{\infty} \subseteq L^p$ and $X_n \xrightarrow{L^p} X \in L^p$. Then, $X_n \xrightarrow{L^q} X$. Indeed, by Theorem 3.6, we have $\{X_n\}_{n=1}^{\infty} \subseteq L^q$ and $X \in L^q$. By Jensen's inequality, as $n \to \infty$,

$$(\mathbb{E}[|X_n - X|^q])^{\frac{p}{q}} \leq (\mathbb{E}[|X_n - X|^p]) \to 0$$

Note that almost-sure convergence and L^p convergence do not imply each other, as seen in the following examples:

Example 4.1 Let $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}(\mathbb{R}), \lambda)$, where λ means the Lebesgue measures. Define

$$X_n(\omega) := n^2 \mathbb{1}_{\left(\frac{1}{n+1}, \frac{1}{n}\right]}(\omega),\ \omega \in [0,1].$$

1. Show that X_n converges almost-surely and find the limit X
2. Does X_n converge to X in L^p, where $p \geq 1$?

Solution.

1. Since for each $\omega \in (0,1]$, there exists n large enough such that $\omega \in (1/n, 1/(n+1)]$. Hence, for any $\omega \in (0,1]$. $X_n(\omega) = 0$ when n is sufficiently large. This implies $X_n \to 0$ a.s.
2. We show that X_n does not converge to $X = 0$ in L^1, which also implies X_n is not convergent to 0 in L^p, $p \geq 1$. Indeed,

$$\mathbb{E}[|X_n|] = n^2 \lambda \left(\left(\frac{1}{n+1}, \frac{1}{n} \right] \right) = n^2 \left(\frac{1}{n} - \frac{1}{n+1} \right) = \frac{n}{n+1} \to 1 \neq 0.$$

□

Example 4.2 Let $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}(\mathbb{R}), \lambda)$. Define

$$X_n(\omega) := \begin{cases} 1, & \text{if } \omega \in \left[\frac{k}{2^m}, \frac{k+1}{2^m} \right]; \\ 0, & \text{otherwise,} \end{cases}$$

where $m = \lfloor \log_2 n \rfloor$ and $k = n - 2^m$.

1. Show that $X_n \xrightarrow{L^2} X$ and find the limit X.
2. Does X_n converge to X a.s.?

Solution.

1. For any $n \geq 1$,

$$\mathbb{E}[|X_n|^2] = 1^2 \lambda \left(\left[\frac{k}{2^m}, \frac{k+1}{2^m} \right] \right) = \frac{1}{2^m} \to 0$$

as $n \to \infty$. Hence, $X_n \xrightarrow{L^2} 0$.

2. Fix $m \in \mathbb{N}$. For $n \in \{2^m, 2^m+1, \ldots, 2^{m+1}-1\}$, the sequence X_n corresponds to the 2^m dyadic intervals of length 2^{-m} given by $[k/2^m, (k+1)/2^m]$, $k = 0, 1, \ldots, 2^m - 1$, which partition the entire interval $\Omega = [0, 1]$. When $n = 2^{m+1}$, the cycle repeats with 2^{m+1} intervals of length $2^{-(m+1)}$ covering $[0, 1]$.
   Therefore, for every $\omega \in [0, 1]$, there exist infinitely many n such that $X_n(\omega) = 1$, and infinitely many n such that $X_n(\omega) = 0$. Hence, the sequence $\{X_n(\omega)\}$ does **not** converge for any $\omega \in [0, 1]$.

$\square$

The following presents a way to prove L^2-convergence of a sequence $\{X_n\}_{n=1}^{\infty}$ to a number $\mu \in \mathbb{R}$.

Proposition 4.2 Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables with $X_n \in L^2$ and $\lim_{n\to\infty} \mathbb{E}[X_n] = \mu$. If $\lim_{n\to\infty} \mathrm{Var}[X_n] = 0$, then $X_n \xrightarrow{L^2} \mu$.

Proof. Note that

$$\begin{aligned}
\mathrm{Var}[X_n] &= \mathbb{E}[(X_n - \mathbb{E}[X_n])^2] \\
&= \mathbb{E}\left[\left((X_n - \mu) + (\mu - \mathbb{E}[X_n])^2\right)\right] \\
&= \mathbb{E}\left[(X_n - \mu)^2\right] + 2(\mu - \mathbb{E}[X_n])(\mathbb{E}[X_n - \mu]) + (\mathbb{E}[X_n] - \mu)^2 \\
&= \mathbb{E}\left[(X_n - \mu)^2\right] - (\mathbb{E}[X_n] - \mu)^2.
\end{aligned}$$

Hence,

$$0 = \lim_{n\to\infty} \mathrm{Var}[X_n] = \lim_{n\to\infty} \left(\mathbb{E}\left[(X_n - \mu)^2\right] - (\mathbb{E}[X_n] - \mu)^2\right) = \lim_{n\to\infty} \mathbb{E}\left[(X_n - \mu)^2\right].$$

Therefore, $X_n \xrightarrow{L^2} \mu$. $\square$

Example 4.3 (**Random walk**) Let $\{\xi_n\}_{n=1}^{\infty}$ be an i.i.d. sequence with $\mathbb{P}(\xi_n = 1) = 1/2 = \mathbb{P}(\xi_n = -1)$. For $n \geq 1$, let

$$X_n := \frac{1}{n} \sum_{k=1}^{n} \xi_k.$$

By the ***strong law of large numbers***, $X_n \to \mathbb{E}[\xi_1] = 0$ a.s. Show that $X_n \xrightarrow{L^2} 0$.

Solution. For $n \geq 1$, it is easy to see that $\mathbb{E}[X_n] = \mathbb{E}[\xi_1] = 0$. Hence, it suffices to show that $\mathrm{Var}[X_n] \to 0$. By the i.i.d. property,

$$\mathrm{Var}[X_n] = \frac{1}{n^2} \sum_{k=1}^{n} \mathrm{Var}[\xi_k] = \frac{n\mathrm{Var}[\xi_1]}{n^2} = \frac{\mathrm{Var}[\xi_1]}{n} = \frac{1}{n} \to 0$$

as $n \to \infty$. Therefore, $X_n \xrightarrow{L^2} 0$. $\square$

The last mode of convergence is the weakest among the three:

Definition 4.3 (Convergence in probability) The sequence of random variables $\{X_n\}_{n=1}^{\infty}$ is said to ***converge in probability*** to X if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

In this case, we write $X_n \xrightarrow{P} X$.

The last result compares the three modes of convergence.

Theorem 4.3

1. $X_n \to X$ a.s. implies $X_n \xrightarrow{P} X$;
2. $X_n \xrightarrow{L^p} X$ implies $X_n \xrightarrow{P} X$;
3. $X_n \to X$ a.s. does NOT imply $X_n \xrightarrow{L^p} X$, and $X_n \xrightarrow{L^p} X$ does NOT imply $X_n \to X$.

The fact that $X_n \xrightarrow{L^p} X$ implies $X_n \xrightarrow{P} X$ is a consequence of Chebyeshev's inequality (exercise). The proof of almost sure convergence implies convergent in probability requires more analysis, readers are referred to Chapter 17 *Probability Essentials* by Jacod and Protter for details.

5 Convergence Theorems of Integrals

In Example 4.1, we have seen that $X_n \to 0$ a.s., but $\lim_{n\to\infty} \mathbb{E}[X_n] = 1 \neq 0$. In general, $X_n \to X$ a.s. alone does not guarantee $\mathbb{E}[X_n] \to \mathbb{E}[X]$. This section presents some important convergence theorems that guarantee the convergence of expected values. Interested readers can find the proofs of these theorems in Appendix B.

Theorem 5.1 (Monotone Convergence Theorem (MCT)) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables such that $X_n \geq 0$, $X_n \leq X_{n+1}$ a.s. for all n, and $X_n \to X$ a.s. Then, $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

Theorem 5.2 (Dominated Convergence Theorem (DCT)) Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables such that $X_n \to X$ a.s., and suppose there exists $Y \in L^1$ such that $|X_n| \leq |Y|$ a.s. for all n. Then, $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

Example 5.1 Let $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$. For each $n > 0$, let $X_n : (\Omega, \mathcal{F}) \to ([0,1], \mathcal{B}([0,1]))$ be a random variable defined by

$$X_n(x) = \frac{\sin(ne^x)}{n}, \;\; x \in [0,1].$$

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

<u>*Solution*</u>. Note that $X_n \to 0$ a.s., and

$$|X_n(x)| = \left|\frac{\sin(ne^x)}{n}\right| \leq \frac{1}{n} \leq 1.$$

Hence, by DCT, we have $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[\lim_{n\to\infty} X_n] = 0$. $\square$

Example 5.2 Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of non-negative random variables. Show that

$$\sum_{n=1}^{\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right].$$

<u>*Solution*</u>. Let $S_n := \sum_{i=1}^{n} X_i$, then $S_n \geq 0$, $S_n \leq S_{n+1}$ for any $n \geq 1$, and $S_n \uparrow S_\infty = \sum_{i=1}^{\infty} X_i$. By MCT, we have

$$\sum_{i=1}^{\infty} \mathbb{E}[X_i] = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}[X_i] = \lim_{n\to\infty} \mathbb{E}[S_n] = \mathbb{E}\left[S_\infty\right] = \mathbb{E}\left[\sum_{i=1}^{\infty} X_i\right].$$

$\square$

Example 5.3 Recall from Example 4.1, we have $X_n \to 0$ a.s., but $\mathbb{E}[X_n] \to 1 \neq 0$. Explain why both the monotone convergence and dominated convergence failed to apply.

<u>*Solution*</u>. MCT: the sequence X_n are non-negative, but not monotonic. Indeed, for any $\omega \in (0, 1/2)$, there exists $n > 1$ such that $X_{n-1}(\omega) = 0$, $X_n(\omega) = n^2$, and $X_n(\omega) = 0$. Therefore, the MCT does not apply.
DCT: DCT requires a single integrable function $Y \in L^1$ such that $|X_n| \leq Y$ a.s. for all n. Since

$$X_n(\omega) = n^2 \quad \text{for } \omega \in \left(\frac{1}{n+1}, \frac{1}{n}\right],$$

the domination condition $|X_n(\omega)| \leq Y(\omega)$ implies $Y(\omega) \geq n^2$ for $\omega \in (1/(n+1), 1/n]$.

Therefore,

$$Y(\omega) \geq \sum_{n=1}^{\infty} n^2 \mathbb{1}_{\left(\frac{1}{n+1}, \frac{1}{n}\right]}(\omega).$$

But the right-hand side defines a function whose integral is

$$\mathbb{E}[Y] = \sum_{n=1}^{\infty} n^2 \left(\frac{1}{n} - \frac{1}{n+1}\right) = \sum_{n=1}^{\infty} \frac{n^2}{n(n+1)} = \sum_{n=1}^{\infty} \frac{n}{n+1} = \infty.$$

Hence, no integrable dominating function Y exists, and the DCT cannot be applied.

□

6 Computations of Expected Values

This section provides some basic rules in computing expectations, most of which have been covered in introductory probability courses.

Theorem 6.1 (Change of variables formula) Let $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ be a random variable with distribution P_X, and $g : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a measurable function such that $\mathbb{E}[|g(X)|] < \infty$. Then,

$$\mathbb{E}[g(X)] = \int_S g(x) dP_X(x).$$

Proof. The formula is proven by the 3-step approach: from simple functions, non-negative functions, to general measurable functions.

Suppose that

$$g(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(x),$$

where $A_1, \ldots, A_n \in \mathcal{S}$, $A_i \cap A_j = \varnothing$ for $i \neq j$. Then,

$$\mathbb{E}[g(X)] = \mathbb{E}\left[\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(X)\right] = \sum_{i=1}^{n} a_i \mathbb{E}[\mathbb{1}_{A_i}(X)] = \int_S \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(x) dP_X(x) = \int_S g(x) dP_X(x).$$

Next, suppose that g is non-negative. Let $\{g_n\}_{n=1}^{\infty}$ be the dyadic approximation of g; see (2). Then, g_n is simple, and $g_n \uparrow g$ a.s. Using the MCT twice, we have

$$\mathbb{E}[g(X)] = \lim_{n\to\infty} \mathbb{E}[g_n(X)] = \lim_{n\to\infty} \int_S g_n(x) dP_X(x) = \int_S g(x) dP_X(x).$$

Note that the MCT was used in the first and the last equality.

Finally, for any measurable function g, we can write $g = g^+ - g^-$. Then,

$$\begin{aligned}
\mathbb{E}[f(X)] &= \mathbb{E}[g^+(X)] - \mathbb{E}[g^-(X)] \\
&= \int_S g^+(x) dP_X(x) - \int_S g^-(x) dP_X(x) \\
&= \int_S [g^+(x) - g^-(x)] dP_X(x) = \int_S g(x) dP_X(x).
\end{aligned}$$

□

Based on the range of values X could take, the expected values can be computed as follows:

1. X can only take countably many values $S = \{x_n\}_{n=1}^{\infty}$:
   - X is said to be a ***discrete random variable***.
   - P_X is characterized by the ***probability mass function*** (pmf), defined by

   $$P_X(x) = \mathbb{P}(X = x), \ x \in S.$$

   - $\mathbb{E}[g(X)]$ can be computed by

   $$\mathbb{E}[g(X)] = \sum_{x \in S} g(x) P_X(x).$$

2. X can take uncountably many values S, and P_X is absolutely continuous with respect to the Lebesgue measure:
   - X is said to be a ***continuous random variable***.
   - The distribution $\mathbb{P}_X$ is characterized by a ***probability density function*** (pdf) f_X, such that for any Borel set $A \subseteq S$,

   $$\mathbb{P}(X \in A) = \int_A f_X(x)\, dx.$$

   - If the distribution function $F_X(x)$ is differentiable at x, then $f_X(x) = \frac{d}{dx} F_X(x)$.
   - $\mathbb{E}[g(X)]$ can be computed by

   $$\mathbb{E}[g(X)] = \int_S g(x) f_X(x)\, dx.$$

Example 6.1 Let X be a *Cauchy* random variable with the following density function:

$$f_X(x) = \frac{1}{\pi} \frac{1}{1 + x^2}, \ x \in \mathbb{R}.$$

1. Show that $\int_{\mathbb{R}} f(x)dx = 1$.
2. Show that $\mathbb{E}[X^+] = \mathbb{E}[X^-] = \infty$, and so $X \notin L^1$.
3. Can we compute $\mathbb{E}[X]$?

Solution.

1. By a direct integration,

$$\int_{\mathbb{R}} f(x)\,dx = \frac{1}{\pi}\int_{\mathbb{R}} \frac{dx}{1+x^2} = \frac{1}{\pi}\tan^{-1}(x)\big|_{-\infty}^{\infty} = \frac{1}{\pi}\left(\frac{\pi}{2} - \frac{-\pi}{2}\right) = 1.$$

2. For $|x| \geq 1$, we have

$$\frac{|x|}{1+x^2} \geq \frac{|x|}{x^2+x^2} = \frac{1}{2x}.$$

Hence,

$$\mathbb{E}[X^+] = \int_0^\infty xf(x)\,dx = \int_0^\infty \frac{x}{1+x^2}dx \geq \int_1^\infty \frac{dx}{2x} = \infty.$$

Likewise, we can show that $\mathbb{E}[X^-] = \infty$. This implies $\mathbb{E}[|X|] = \mathbb{E}[X^+] + \mathbb{E}[X^-] = \infty$, and thus $X \notin L^1$.

3. Despite the integrand $x/(1+x^2)$ is an odd function, $\mathbb{E}[X] \neq 0$, since both $\mathbb{E}[X^+], \mathbb{E}[X^-]$ are not finite. In this case, $\mathbb{E}[X]$ is not well-defined.

□

Example 6.2 The ***moment generating function*** of the random variable X is defined as

$$M_X(t) = \mathbb{E}[e^{tX}],\ t \in \mathbb{R},$$

provided that $e^{tX} \in L^1$. Let X be a standard normal variable, $X \sim \mathcal{N}(0,1)$, which admits the following probability density function

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

Compute $M_X(t)$.

Solution. Using the change of variables formula,

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \int_{-\infty}^{\infty} e^{tx}\varphi(x)\,dx \\ &= \int_{-\infty}^{\infty} e^{tx}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right)dx \\ &= e^{\frac{t^2}{2}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-t)^2}{2}}\,dx = e^{\frac{t^2}{2}}. \end{aligned}$$

□

Appendices

A Proofs of Useful Inequalities

A.1 Proof of Hölder's Inequality

The inequality clearly holds if $p = 1, q = \infty$ (or when $q = 1, p = \infty$), since

$$\mathbb{E}[|XY|] \leq \|Y\|_\infty \mathbb{E}[|X|] = \|X\|_1 \|Y\|_\infty.$$

If $\|X\|_p = 0$ or $\|Y\|_q = 0$, we must have $|XY| = 0$ a.s., and thus the inequality also holds. Therefore, it suffices to consider $p, q \in (1, \infty)$, and $\|X\|_p, \|Y\|_q > 0$.

Using Young's inequality, for any $x, y \geq 0$ and $p, q \in (1, \infty)$ such that $1/p + 1/q = 1$, we have

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Using this, and by taking

$$x = \frac{|X|}{\|X\|_p} \text{ and } y = \frac{|Y|}{\|Y\|_q},$$

we have

$$|XY| \leq \|X\|_p \|Y\|_q \left(\frac{|X|^p}{p\|X\|_p^p} + \frac{|Y|^q}{q\|Y\|_q^q} \right)$$

The desired inequality follows by taking expectations on both sides.

A.2 Proof of Jensen's Inequality

Let $\mu := \mathbb{E}[X]$. By the convexity of φ, we can always find $a \in \mathbb{R}$ such that the linear function $l(x) := a(x - \mu) + \varphi(\mu)$ satisfies $l(x) \leq \varphi(x)$ for all $x \in \mathbb{R}$. Using this fact, we have

$$\mathbb{E}[\varphi(X)] \geq \mathbb{E}[l(X)] = a\mathbb{E}[X - \mu] + \varphi(\mu) = \varphi(\mathbb{E}[X]).$$

B Proof of MCT and DCT

We provide the proofs of the MCT and DCT. To this end, we introduce the following key lemma:

Theorem B.1 (Fatou's Lemma) Let $\{X_n\}_{n=1}^\infty$ be a sequence of non-negative measurable functions. Then,

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n].$$

B.1 Proof of MCT

We prove the MCT using Fatou's lemma. Since $X_n \uparrow X$, we have $\liminf_{n\to\infty} X_n = X$ a.s. The monotonicity also implies $\mathbb{E}[X_n] \uparrow \mathbb{E}[X]$, whence $\liminf_{n\to\infty} \mathbb{E}[X_n] = \lim_{n\to\infty} \mathbb{E}[X_n]$. Using the fact that $X_n \geq 0$ a.s., we can apply Fatou's lemma to conclude that

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

This proves one side of the convergence theorem.

To complete the proof, we need to show that $\mathbb{E}[X] \geq \lim_{n\to\infty} \mathbb{E}[X_n]$. To this end, consider $Y_n := X - X_n$. Then $Y_n \geq 0$ a.s. for all $n \geq 1$, $\liminf_{n\to\infty} Y_n = 0$ a.s., and $\liminf_{n\to\infty} \mathbb{E}[Y_n] = \liminf_{n\to\infty}(\mathbb{E}[X] - \mathbb{E}[X_n]) = \mathbb{E}[X] - \lim_{n\to\infty} \mathbb{E}[X_n]$. By Fatou's lemma,

$$0 = \mathbb{E}\left[\lim_{n\to\infty} Y_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[Y_n] = \mathbb{E}[X] - \lim_{n\to\infty} \mathbb{E}[X_n],$$

which proves the desired inequality.

B.2 Proof of DCT

Similar to the proof of MCT, we establish DCT using Fatou's lemma. Since $|X_n| \leq Y$ and $X_n \to X$ a.s., we also have $|X| = \lim |X_n| \leq Y$ a.s. Hence, $X \in L^1$ by comparison with Y.

By applying Fatou's lemma to the non-negative functions $X_n + Y$, we have

$$\mathbb{E}[X + Y] \leq \liminf_{n\to\infty} \mathbb{E}[X_n + Y] = \liminf_{n\to\infty}(\mathbb{E}[X_n] + \mathbb{E}[Y]),$$

so

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

Next, apply Fatou's Lemma to $-X_n + Y$, which are also non-negative:

$$\mathbb{E}[-X + Y] \leq \liminf_{n\to\infty} \mathbb{E}[-X_n + Y] = \liminf_{n\to\infty}(-\mathbb{E}[X_n] + \mathbb{E}[Y]),$$

which gives

$$-\mathbb{E}[X] \leq \liminf_{n\to\infty}(-\mathbb{E}[X_n]), \quad \text{or} \quad \mathbb{E}[X] \geq \limsup_{n\to\infty} \mathbb{E}[X_n].$$

Therefore,

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

which implies

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