THE MARTINGALE CONVERGENCE THEOREM: A MEASURE-THEORETIC PROOF AND APPLICATIONS

ISHAAN SHARMA

Abstract. The Martingale Convergence Theorem (MCT) is a central result in probability theory, asserting that any L^1-bounded martingale converges almost surely to an integrable limit. This paper presents a self-contained proof of the MCT, beginning with core measure-theoretic foundations such as σ-algebras, measures, and Lebesgue integration, and developing the probabilistic framework of random variables, conditional expectation, and martingales. Key tools including Fatou's Lemma, the Monotone Convergence Theorem, and Doob's Upcrossing Lemma are established and applied to prove the theorem. Applications and counterexamples are explored to illustrate the theorem's scope and limitations. We conclude with extensions to continuous-time processes and L^p-convergence, and include a detailed application in mathematical finance as well as in the proof of the SLLN.

Introduction

Martingales are a central concept in modern probability theory, with applications ranging from stochastic calculus and mathematical finance to statistical estimation, potential theory, and online algorithms. A martingale represents a "fair game", where the expected future value, conditioned on the present, equals the current value. The Martingale Convergence Theorem (MCT) reveals a deep regularity in such processes: under a simple L^1-boundedness condition, martingales converge almost surely to a well-defined random variable.

0.1. Statement of the Main Theorem.

Theorem 0.1 (Martingale Convergence Theorem)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be an L^1-bounded martingale adapted to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$. That is,*

$$\sup_n \mathbb{E}[|X_n|] < \infty.$$

Then there exists an integrable random variable $X_\infty \in L^1$ such that

$$X_n \xrightarrow{a.s.} X_\infty.$$

Moreover, if (X_n) is uniformly integrable, the convergence also holds in L^1.

0.2. **Outline of the Paper.** This paper is structured to provide a clear and rigorous derivation of the Martingale Convergence Theorem from first principles. The main components are as follows:

- Section 1 introduces the foundational elements of measure theory including σ-algebras, measures, and measurable spaces.
- In Section 2, we develop the theory and purpose of Lebesgue integration, and discuss L^p spaces in probability.

Date: July 1, 2025.

- Section 3 introduces probability spaces and random variables, formalizing the probabilistic framework as a special case of measure theory.
- In Section 4, we define expectation as a Lebesgue integral and frame conditional probability and expectation as random variables.
- Section 5 presents the fundamental properties of conditional expectation: uniqueness, linearity, the tower property, positivity, measurability, and Jensen's inequality.
- Section 6 defines filtrations, martingales, submartingales, and supermartingales, with motivating examples such as symmetric and biased random walks.
- Section 7 presents the Monotone Convergence Theorem and Fatou's Lemma, which are essential for managing limits of integrals throughout the convergence proof.
- In Section 8, we prove Doob's Upcrossing Lemma and its corollaries, which control oscillations in submartingales and form the heart of the convergence argument.
- Section 9 presents a detailed, step-by-step proof of the Martingale Convergence Theorem, emphasizing the role of upcrossings and boundedness.
- Section 10 explores extensions and applications: convergence in L^p, continuous-time martingales, an example from mathematical finance in derivative valuation, and an elegant proof of the SLLN using the MCT.
- Finally, Section 11 summarizes the main results and discusses the theorem as a whole.

Each section is self-contained and motivated by examples, with the goal of making the Martingale Convergence Theorem accessible to students and researchers.

1. Measure Theory Foundations

1.1. **Sigma-Algebras and Measurable Spaces.** Measure theory provides the mathematical framework for modern probability theory by generalizing concepts like length, area, and volume to abstract spaces. Probability measures are special cases of measures, and integration with respect to them forms the backbone of expectations, variances, and distributions.

Definition 1.1. A *σ-algebra* $\mathcal{F}$ on a set Ω is a collection of subsets of Ω satisfying:

(1) $\Omega \in \mathcal{F}$,
(2) If $A \in \mathcal{F}$, then $A^c = \Omega \setminus A \in \mathcal{F}$ (closed under complements),
(3) If $(A_n)_{n=1}^{\infty} \subseteq \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ (closed under countable unions).

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

A σ-algebra defines the class of sets for which we can consistently assign a measure or probability. Closure under complements and countable unions ensures compatibility with limits and allows for the application of convergence theorems central to probability theory.

Example (Finite σ-algebra). Let $\Omega = \{1, 2, 3\}$. Then the power set $\mathcal{P}(\Omega)$ is a σ-algebra. A smaller σ-algebra could be

$$\mathcal{F} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\},$$

which is closed under complements and countable unions.

Example (Borel σ-algebra). On $\mathbb{R}$, the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is generated by open intervals. It includes all open and closed sets, countable intersections and unions of such sets, and much more. It is the smallest σ-algebra containing all open sets.

1.2. **Measures and Measure Spaces.**

Definition 1.2. A *measure* μ on a measurable space $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ satisfying:

(1) $\mu(\emptyset) = 0$,
(2) (Countable additivity) For any sequence of pairwise disjoint sets $(A_n)_{n=1}^{\infty} \subseteq \mathcal{F}$,

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

The triple $(\Omega, \mathcal{F}, \mu)$ is called a *measure space*.

Example (Lebesgue measure). The Lebesgue measure λ on $\mathbb{R}$ assigns to each interval (a, b) the length $b-a$, and extends this to the Borel sets via countable additivity. It underpins the Lebesgue integral, which generalizes the Riemann integral to a far broader class of functions.

Example (Counting measure). Let $\Omega = \mathbb{N}$ and define $\mu(A) = |A|$ for all $A \subseteq \mathbb{N}$, where $|A|$ denotes the (possibly infinite) cardinality. Then μ is a measure, called the *counting measure*. This is useful for defining discrete probability spaces and sums as integrals.

2. Measurable Functions and Integration

2.1. **Measurable Functions.**

Before we define integration on a probability space, we must identify the class of functions that are integrable: measurable functions.

Definition 2.1 (Measurable Function)**.** Let $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$ be measurable spaces. A function $X : \Omega \to S$ is said to be $\mathcal{F}$*-measurable* (or simply measurable) if for every set $A \in \mathcal{S}$, the preimage $X^{-1}(A) \in \mathcal{F}$.

In this paper, we will primarily be concerned with real-valued functions $X : \Omega \to \mathbb{R}$, where $\mathbb{R}$ is equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{R})$. In this setting, the following definition is more accurate.

Definition 2.2. Let $(\Omega, \mathcal{F})$ be a measurable space. A function $X : \Omega \to \mathbb{R}$ is said to be $\mathcal{F}$*-measurable* if for every Borel set $B \subseteq \mathbb{R}$, the preimage $X^{-1}(B) \in \mathcal{F}$.

Key Properties of Measurable Functions.

- Any continuous function from a topological space to $\mathbb{R}$ is Borel measurable.
- The pointwise limit of a sequence of measurable functions is measurable.
- The composition of measurable functions is measurable.

Why Measurability Matters. Measurability ensures that the preimages of "observable" events (in $\mathbb{R}$) correspond to measurable sets in Ω, making integration well-defined. Only measurable functions can be Lebesgue-integrated.

2.2. **Simple Functions and the Lebesgue Integral.**

We begin with the class of functions on which the integral is first defined: simple functions.

Definition 2.3 (Simple Function)**.** Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\varphi : \Omega \to \mathbb{R}$ is called a *simple function* if it takes on only finitely many real values, i.e.,

$$\varphi(\omega) = \sum_{i=1}^{n} a_i \, \mathbf{1}_{A_i}(\omega),$$

where $a_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$ are pairwise disjoint measurable sets, and $\bigcup_{i=1}^{n} A_i = \Omega$.

Definition 2.4 (Integral of a Non-negative Simple Function)**.** Let φ be a non-negative simple function on a measure space $(\Omega, \mathcal{F}, \mu)$, i.e., $a_i \geq 0$ for all i. Then the *Lebesgue integral* of φ with respect to μ is defined by:

$$\int_\Omega \varphi \, d\mu = \sum_{i=1}^{n} a_i \, \mu(A_i).$$

This definition is well-defined because the representation of φ can be reduced to a unique (up to null sets) canonical form with disjoint sets.

Definition 2.5 (Integral of a Non-negative Measurable Function)**.** Let $f : \Omega \to [0, \infty]$ be a non-negative measurable function. The *Lebesgue integral* of f with respect to μ is defined by:

$$\int_\Omega f \, d\mu = \sup \left\{ \int_\Omega \varphi \, d\mu \, : \, 0 \leq \varphi \leq f, \, \varphi \text{ simple} \right\}.$$

This definition extends the integral from simple functions to arbitrary non-negative measurable functions via supremum approximation from below.

Definition 2.6 (Integral of a General Measurable Function)**.** Let $f : \Omega \to \mathbb{R}$ be an arbitrary measurable function. Decompose it into its positive and negative parts:

$$f^+ = \max(f, 0), \qquad f^- = \max(-f, 0),$$

so that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Then, if $\int f^+ \, d\mu < \infty$ and $\int f^- \, d\mu < \infty$, we define:

$$\int_\Omega f \, d\mu = \int_\Omega f^+ \, d\mu - \int_\Omega f^- \, d\mu.$$

If either of these integrals is infinite, we say the integral of f is undefined (or diverges).

This construction ensures that the Lebesgue integral handles signed functions correctly and aligns with our intuitive notion of area under a curve (positive minus negative contribution).

Remark 2.7 (Comparison with Riemann Integration). Unlike the Riemann integral, the Lebesgue integral is defined with respect to a measure rather than partitioning the domain into intervals. This allows it to integrate functions with dense sets of discontinuities, manage changes on sets of measure zero and handle limit operations more robustly.

Example. Let $f = \mathbf{1}_{\mathbb{Q} \cap [0,1]}$, the indicator function of the rational numbers in the interval $[0, 1]$. That is,

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q} \cap [0, 1], \\ 0, & x \in [0, 1] \setminus \mathbb{Q}. \end{cases}$$

This function is measurable because the set $\mathbb{Q} \cap [0, 1]$ is countable and hence a Borel set. However, f is discontinuous at every point of the interval $[0, 1]$, since every open interval contains both rational and irrational numbers. Consequently, f is not Riemann integrable, as it is discontinuous on a set of positive measure. Nonetheless, the Lebesgue integral of f over $[0, 1]$ exists and is equal to the measure of the set where f is nonzero:

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

because the rationals form a countable set and thus have Lebesgue measure zero.

2.3. L^p **Spaces and Integrability.** A central concept in both measure theory and probability is that of integrability. The integrability of a function determines whether its expected value or more general integral can be meaningfully defined.

Definition 2.8. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. A measurable function $X : \Omega \to \mathbb{R}$ is said to be *(Lebesgue) integrable* with respect to μ if

$$\int_\Omega |X| \, d\mu < \infty.$$

The set of all such functions is denoted by $L^1(\mu)$, the space of absolutely integrable functions.

The L^p Spaces. More generally, for $1 \le p < \infty$, we define the L^p-spaces:

$$L^p(\mu) = \left\{ X : \Omega \to \mathbb{R} \text{ measurable} \;\middle|\; \int_\Omega |X|^p \, d\mu < \infty \right\}.$$

These are normed vector spaces under the norm

$$\|X\|_p = \left(\int_\Omega |X|^p \, d\mu \right)^{1/p}.$$

The space $L^\infty(\mu)$ consists of essentially bounded measurable functions.

Example. Let $f(x) = x^{-\alpha}$ on the interval $(0, 1]$, and consider the measure space $([0, 1], \mathcal{B}, \lambda)$, where λ is the Lebesgue measure. Then

$$\int_0^1 |f(x)|^p \, dx = \int_0^1 x^{-\alpha p} \, dx.$$

This integral converges if and only if $\alpha p < 1$. Therefore,

$$f \in L^p([0, 1]) \quad \Longleftrightarrow \quad \alpha < \frac{1}{p}.$$

2.4. **Why Lebesgue Integration in Probability Theory?** In probability, we frequently work with expectations (i.e., integrals), limits of random variables, and conditional distributions. The Lebesgue integral is preferred over the Riemann integral because:

- It accommodates limits and infinite sequences of functions rigorously.
- It is defined on abstract measure spaces, enabling generality in probabilistic models.
- It ignores events of probability zero — aligning with the probabilistic notion that such events do not affect outcomes.
- It generalizes the Riemann integral in the sense that every Riemann integrable function is also Lebesgue integrable, but the converse does not hold.

Lebesgue integration provides the rigorous analytical tools needed for advanced probability theory, including martingale convergence, where control of expectations and limiting behavior is crucial. More information can be found in [RF10, Axl22].

3. Probability Theory as Measure Theory

3.1. **Probability Spaces.** Probability theory is naturally framed within the language of measure theory [Bil95, Dur19]. This foundational connection, formalised by Kolmogorov [Kol56], allows probabilistic concepts to be rigorously interpreted as special cases of measure-theoretic constructions. In particular, probability measures are measures of total mass one.

Definition 3.1. A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where:

- Ω is the *sample space*, the set of all possible outcomes;
- $\mathcal{F}$ is a σ-algebra of subsets of Ω, called the collection of *events*;
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a *probability measure*, satisfying $\mathbb{P}(\Omega) = 1$.

Unlike general measures, which may assign infinite total mass and are not necessarily bounded, probability measures are normalized so that the entire space has measure one. This reflects the interpretation of $\mathbb{P}(A)$ as the likelihood of event A occurring.

Example (Coin Toss Space). For an infinite sequence of coin tosses, take $\Omega = \{H, T\}^{\mathbb{N}}$. Let $\mathcal{F}$ be the σ-algebra generated by finite cylinder sets, and $\mathbb{P}$ the product measure with $\mathbb{P}(H) = \mathbb{P}(T) = 1/2$. This models fair, independent coin tosses.

3.2. **Random Variables.**

Definition 3.2. A *random variable* is a measurable function $X : \Omega \to \mathbb{R}$. That is, for every Borel set $B \subseteq \mathbb{R}$, the preimage $X^{-1}(B) \in \mathcal{F}$.

Remark 3.3. Random variables are not inherently random; rather, they are deterministic functions from the sample space to the reals. Their randomness arises from the underlying uncertainty described by the probability measure $\mathbb{P}$.

Example. Let X be a standard normal random variable. Then $\Omega = \mathbb{R}$, $\mathcal{F}$ is the Borel σ-algebra, and for any Borel set A,

$$\mathbb{P}(A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2}\, dx.$$

3.3. **Expectation.** Measure theory provides a general and robust theory of integration via the Lebesgue Integral, which extends the classical Riemann integral and overcomes many of its limitations—particularly in the context of convergence and measurability.

Definition 3.4. Let $X : \Omega \to \mathbb{R}$ be a measurable function on a measure space $(\Omega, \mathcal{F}, \mu)$. The *Lebesgue integral* of X with respect to μ is denoted by

$$\int_\Omega X\, d\mu.$$

In the context of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the Lebesgue integral of a random variable X is referred to as its *expectation*:

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

Example. Let $X(\omega) = \omega$ on $[0, 1]$ with $\mathbb{P}$ the Lebesgue measure. Then the expectation of X is

$$\mathbb{E}[X] = \int_0^1 x\, dx = \frac{1}{2}.$$

Example. Let X be the indicator function of a coin flip resulting in heads: $X(H) = 1$, $X(T) = 0$. Then for $\mathbb{P}(H) = p$, the expected value is

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p.$$

3.4. **Stochastic Processes.** A *stochastic process* is a mathematical model for a system that evolves randomly over time and appear in diverse areas such as physics, finance, and biology. Formally, it is defined as a collection of random variables indexed by a parameter that typically represents time.

Definition 3.5 (Stochastic Process)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *stochastic process* is a family of random variables $(X_t)_{t\in T}$, where each $X_t : \Omega \to \mathbb{R}$ is measurable with respect to $\mathcal{F}$. The index set T typically represents time and may be discrete ($T = \mathbb{N}$) or continuous ($T = [0, \infty)$).

That is, a stochastic process is a function:

$$X : T \times \Omega \to \mathbb{R}, \quad (t, \omega) \mapsto X_t(\omega),$$

such that for each fixed $t \in T$, the map $\omega \mapsto X_t(\omega)$ is a random variable.

Intuitively, for each outcome $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ describes a sample path or realization of the process, capturing how the system evolves over time for that particular outcome. Conversely, for a fixed time $t \in T$, the random variable X_t represents the state of the process at that time.

Example (Simple Random Walk). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which is defined a sequence of i.i.d. random variables $(\xi_n)_{n\in\mathbb{N}}$ taking values in $\{-1, +1\}$ with equal probability $1/2$. The process $(S_n)_{n\in\mathbb{N}}$ defined by

$$S_0 = 0, \quad S_n = \sum_{k=1}^{n} \xi_k$$

is called a *Simple Symmetric Random Walk*. It models the position of a particle moving on $\mathbb{Z}$ that takes a step left or right with equal probability at each time step.

4. Conditional Probability and Expectation

4.1. **Conditional Probability.** In elementary probability, conditional probability is defined via the ratio:

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

for events $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$. This formula, however, fails to generalize in measure theory. Notably, if $\mathbb{P}(B) = 0$, the definition is meaningless, yet such events often arise in continuous probability spaces. A more robust formulation uses conditional probability with respect to a σ-algebra [Dur19, Kle13].

Example. Consider the product probability space $([0, 1]^2, \mathcal{B}([0, 1]^2), \lambda)$, where λ is Lebesgue measure. Let $X(x, y) = x$ be the first coordinate projection and define the event,

$$A = \{(x, y) \in [0, 1]^2 : y \leq x\}.$$

We would like to compute the probability $\mathbb{P}(A \mid X)$. However, for a fixed $x \in [0, 1]$, the vertical line $\{x\} \times [0, 1]$ has measure zero in λ and thus, the basic definition of conditional probability as a ratio is undefined. In this case, the conditional probability is given by:

$$\mathbb{P}(A \mid X)(x) = \lambda_1(A_x) = x.$$

Definition 4.1 (Conditional Probability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. For any event $A \in \mathcal{F}$, the conditional probability of A given $\mathcal{G}$, denoted $\mathbb{P}(A \mid \mathcal{G})$, is defined as a $\mathcal{G}$-measurable function $\mathbb{P}(A \mid \mathcal{G}) : \Omega \to [0, 1]$ satisfying

$$\int_B \mathbb{P}(A \mid \mathcal{G}) \, d\mathbb{P} = \mathbb{P}(A \cap B) \quad \text{for all } B \in \mathcal{G}.$$

This means that $\mathbb{P}(A \mid \mathcal{G})$ is itself a random variable, not just a number. It is measurable with respect to $\mathcal{G}$, and hence can be interpreted as the probability of A under the information encoded by $\mathcal{G}$.

4.2. **Conditional Expectation.**

Definition 4.2 (Conditional Expectation)**.** Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. The *conditional expectation* of X given $\mathcal{G}$, denoted $\mathbb{E}[X \mid \mathcal{G}]$, is a $\mathcal{G}$-measurable random variable such that:

(1) $\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable,
(2) For all $A \in \mathcal{G}$,

$$\int_A \mathbb{E}[X \mid \mathcal{G}] \, d\mathbb{P} = \int_A X \, d\mathbb{P}.$$

As with conditional probability, the conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ is a random variable itself—specifically, a function measurable with respect to the information encoded in $\mathcal{G}$. It represents the best approximation of X given the information available in $\mathcal{G}$.

Example (Discrete conditioning). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space where we first flip a fair coin, then roll a fair six-sided die. Let the random variable X be the die roll result, and let $\mathcal{G} = \sigma(C)$, where C denotes the coin outcome (either H or T). Since the coin and die are independent, and the die is uniform,

$$\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X] = 3.5 \quad \text{a.s.}$$

In this case, knowing the coin flip gives no information about the die result, so the conditional expectation equals the marginal expectation.

Example (Conditional Expectation under Symmetry). Let $(X, Y) \sim \text{Uniform}([0, 1]^2)$, and define $Z = X + Y$. Let $\mathcal{G} = \sigma(Z)$. To compute $\mathbb{E}[X \mid \mathcal{G}]$, we observe that the line $X + Y = z$ intersects the unit square in the interval

$$X \in [\max(0, z - 1), \min(1, z)] \, .$$

Given $Z = z$, the conditional distribution of X is uniform over this interval. Thus,

$$\mathbb{E}[X \mid Z = z] = \frac{1}{2} \left(\max(0, z - 1) + \min(1, z) \right).$$

Since this is a function of Z, we conclude $\mathbb{E}[X \mid \mathcal{G}] = f(Z)$, which is $\mathcal{G}$-measurable.

5. Properties of Conditional Expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Unless stated otherwise, all random variables are assumed to be integrable (i.e., belong to $L^1(\Omega, \mathcal{F}, \mathbb{P})$). The conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ plays the role of a best $\mathcal{G}$-measurable approximation of X, and it satisfies the following foundational properties.

Uniqueness. *Conditional expectation is defined up to sets of measure zero. That is, any two versions of* $\mathbb{E}[X \mid \mathcal{G}]$ *are equal almost surely.*

Proposition 5.1 (Uniqueness)**.** *Let* $X \in L^1$*, and let* Z_1, Z_2 *be two* $\mathcal{G}$*-measurable functions such that*

$$\int_A Z_1 \, d\mathbb{P} = \int_A X \, d\mathbb{P} = \int_A Z_2 \, d\mathbb{P}, \quad \text{for all } A \in \mathcal{G}.$$

Then $Z_1 = Z_2$ *almost surely.*

Proof. Let $D = \{\omega \in \Omega : Z_1(\omega) > Z_2(\omega)\} \in \mathcal{G}$. Then,

$$\int_D (Z_1 - Z_2) \, d\mathbb{P} = \int_D Z_1 \, d\mathbb{P} - \int_D Z_2 \, d\mathbb{P} = \int_D X \, d\mathbb{P} - \int_D X \, d\mathbb{P} = 0.$$

But $Z_1 - Z_2 > 0$ on D, so this implies $\mathbb{P}(D) = 0$. Similarly, $\mathbb{P}(Z_2 > Z_1) = \mathbb{P}(D') = 0$, and thus $Z_1 = Z_2$ almost surely. □

Linearity. *Conditional expectation behaves like a linear operator.*

Proposition 5.2 (Linearity)**.** *For* $a, b \in \mathbb{R}$ *and* $X, Y \in L^1$*,*

$$\mathbb{E}[aX + bY \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}] \quad a.s.$$

Proof. Let $Z = aX + bY$. For any $A \in \mathcal{G}$,

$$\begin{aligned} \int_A \mathbb{E}[Z \mid \mathcal{G}] \, d\mathbb{P} &= \int_A Z \, d\mathbb{P} = a \int_A X \, d\mathbb{P} + b \int_A Y \, d\mathbb{P} \\ &= \int_A (a\mathbb{E}[X \mid \mathcal{G}] + b\mathbb{E}[Y \mid \mathcal{G}]) \, d\mathbb{P}. \end{aligned}$$

By uniqueness, the two sides are equal almost surely. □

Tower Property (Iterated Expectations). *Conditioning twice is the same as conditioning once on the smaller* σ*-algebra. This is known as the tower property.*

Proposition 5.3 (Tower Property)**.** *Let* $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$*. For* $X \in L^1$*,*

$$\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}\right] = \mathbb{E}[X \mid \mathcal{H}] \quad a.s.$$

Proof. Let $A \in \mathcal{H}$. Then:

$$\int_A \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] \, d\mathbb{P} = \int_A \mathbb{E}[X \mid \mathcal{G}] \, d\mathbb{P} = \int_A X \, d\mathbb{P} = \int_A \mathbb{E}[X \mid \mathcal{H}] \, d\mathbb{P}.$$

Since both sides are $\mathcal{H}$-measurable, they are equal almost surely. □

Positivity. *Conditional expectation preserves non-negativity.*

Proposition 5.4 (Positivity)**.** *If* $X \geq 0$ *almost surely, then* $\mathbb{E}[X \mid \mathcal{G}] \geq 0$ *almost surely.*

Proof. Suppose $\mathbb{E}[X \mid \mathcal{G}] < 0$ on a set $A \in \mathcal{G}$ with positive measure. Then:

$$\int_A \mathbb{E}[X \mid \mathcal{G}] \, d\mathbb{P} < 0,$$

contradicting the fact that

$$\int_A \mathbb{E}[X \mid \mathcal{G}] \, d\mathbb{P} = \int_A X \, d\mathbb{P} \geq 0,$$

since $X \geq 0$ a.s. Therefore, $\mathbb{E}[X \mid \mathcal{G}] \geq 0$ a.s. □

Factoring Out Measurable Random Variables. *Measurable functions can be pulled out of the conditional expectation.*

Proposition 5.5 (Measurable Multipliers)**.** *Let $X \in L^1$, and let Y be a bounded, $\mathcal{G}$-measurable random variable. Then:*

$$\mathbb{E}[XY \mid \mathcal{G}] = Y \cdot \mathbb{E}[X \mid \mathcal{G}] \quad a.s.$$

Proof. For any $A \in \mathcal{G}$, since Y is $\mathcal{G}$-measurable:

$$\int_A XY \, d\mathbb{P} = \int_A Y \cdot X \, d\mathbb{P} = \int_A Y \cdot \mathbb{E}[X \mid \mathcal{G}] \, d\mathbb{P} = \int_A (Y \cdot \mathbb{E}[X \mid \mathcal{G}]) \, d\mathbb{P}.$$

By uniqueness, the result follows. □

Jensen's Inequality. *Convex functions commute with conditional expectations in an inequality.*

Theorem 5.6 (Conditional Jensen's Inequality)**.** *Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex (and thus Borel-measurable) function such that $\varphi(X) \in L^1$. Then:*

$$\varphi\left(\mathbb{E}[X \mid \mathcal{G}]\right) \leq \mathbb{E}\left[\varphi(X) \mid \mathcal{G}\right] \quad a.s.$$

Proof. By convexity of φ, for any $x_0 \in \mathbb{R}$, there exists an affine support:

$$\varphi(x) \geq \varphi(x_0) + a(x_0)(x - x_0) \quad \forall\, x \in \mathbb{R}.$$

Setting $x_0 = \mathbb{E}[X \mid \mathcal{G}]$ and $x = X$, we take conditional expectations to get,

$$\mathbb{E}[\varphi(X) \mid \mathcal{G}] \geq \varphi(\mathbb{E}[X \mid \mathcal{G}]) + a(\mathbb{E}[X \mid \mathcal{G}]) \cdot \mathbb{E}[X - \mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{G}].$$

Finally, using linearity and the tower property, the inequality simplifies to

$$\varphi\left(\mathbb{E}[X \mid \mathcal{G}]\right) \leq \mathbb{E}\left[\varphi(X) \mid \mathcal{G}\right]$$

as desired. □

6. Martingales

Martingales are fundamental objects in modern probability theory, capturing the idea of a "fair game" in a rigorous, measure-theoretic framework. They are widely used in stochastic processes, statistics, mathematical finance, and the analysis of algorithms [Wil91].

6.1. **Filtrations and Adaptation.** To model the evolution of information over time, we introduce the concept of a filtration.

Definition 6.1 (Filtration)**.** A *filtration* is an increasing sequence $\{\mathcal{F}_n\}_{n\geq 0}$ of sub-σ-algebras of $\mathcal{F}$:

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n \subseteq \mathcal{F}.$$

Intuitively, $\mathcal{F}_n$ represents the information available up to time n.

Example (Filtration for Coin Tosses). Consider a probability space where we repeatedly toss a fair coin. Define $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$, where $Y_k \in \{0, 1\}$ encodes the outcome of the kth toss. Then $\{\mathcal{F}_n\}$ forms a natural filtration representing information revealed after each toss.

6.2. **Definition of Martingales.**

Definition 6.2 (Martingale)**.** Let $\{X_n\}_{n\geq 0}$ be a sequence of integrable random variables adapted to a filtration $\{\mathcal{F}_n\}$. Then $\{X_n\}$ is called a *martingale* if, for all $n \geq 0$,

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n \quad \text{a.s.}$$

Definition 6.3 (Submartingale and Supermartingale)**.** Let $\{X_n\}_{n\geq 0}$ be a sequence of integrable random variables adapted to a filtration $\{\mathcal{F}_n\}$. Then:

- $\{X_n\}$ is a *submartingale* if, for all $n \geq 0$,
$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \geq X_n \quad \text{a.s.}$$
- $\{X_n\}$ is a *supermartingale* if, for all $n \geq 0$,
$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \leq X_n \quad \text{a.s.}$$

Remark 6.4. Martingales model "fair games" where, given the current information, the expected value of the next observation equals the present - no expected gain. Submartingales and supermartingales correspond to "favorable" and "unfavorable" games, respectively, in the sense of expected value trends.

Example (Random Walks as Martingales and Sub/Supermartingales)*.* Let $\{Y_n\}$ be i.i.d. random variables with $\mathbb{P}(Y_n = 1) = p$ and $\mathbb{P}(Y_n = -1) = 1 - p$. Define the partial sums $X_n = \sum_{k=1}^n Y_k$, with filtration $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$. Then:

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n + \mathbb{E}[Y_{n+1}] = X_n + (2p - 1).$$

Consequently:

- When $p = \frac{1}{2}$, $\{X_n\}$ is a martingale (fair game)
- When $p > \frac{1}{2}$, $\{X_n\}$ is a submartingale (favorable game)
- When $p < \frac{1}{2}$, $\{X_n\}$ is a supermartingale (unfavorable game)

Example (Doob Martingale)*.* Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{\mathcal{F}_n\}$ be a filtration. Define

$$X_n = \mathbb{E}[X \mid \mathcal{F}_n].$$

Then $\{X_n\}$ is a martingale. This is known as the Doob martingale associated with X. It reflects the best estimate of X given the information available at time n. The martingale property

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = \mathbb{E}[X \mid \mathcal{F}_n] = X_n$$

follows from the tower property.

Example. Let $(X_n)_{n\geq 0}$ be a submartingale adapted to a filtration $(\mathcal{F}_n)$, and fix a threshold $a \in \mathbb{R}$. Define the process

$$Y_n = (X_n - a)^+ = \max(X_n - a, 0).$$

Then (Y_n) is also a submartingale. This will be used in the proof of the Martingale Convergence Theorem and follows from the convexity of $f(x) = (x-a)^+$. By Jensen's Inequality 5.6 the submartingale property is preserved under convex transformations:

$$\mathbb{E}[Y_n \mid \mathcal{F}_{n-1}] = \mathbb{E}[(X_n - a)^+ \mid \mathcal{F}_{n-1}] \geq (\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] - a)^+.$$

From the submartingale property of X_n, we get:

$$(\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] - a)^+ \geq (X_{n-1} - a)^+ = Y_{n-1}.$$

Combining the two inequalities proves that (Y_n) is a submartingale.

7. Measure-Theoretic Convergence Theorems

Before we approach martingale convergence, we collect two essential tools in integration theory: the Monotone Convergence Theorem and Fatou's Lemma. Both theorems concern the interplay between limits and integration and give sufficient conditions under which these operations commute. Deeper analysis is in [RF10, Axl22].

7.1. **Monotone Convergence Theorem.**

Theorem 7.1 (Monotone Convergence Theorem)**.** *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of non-negative measurable functions such that:*

(1) $f_n \le f_{n+1}$ *for all n (i.e., the sequence is pointwise increasing),*
(2) $f = \lim_{n\to\infty} f_n$ *exists pointwise.*

Then,

$$\lim_{n\to\infty} \mathbb{E}[f_n] = \mathbb{E}[f].$$

Proof. Since $f_n \uparrow f$, we may approximate f from below using simple functions. Let ϕ be any simple function such that $0 \le \phi \le f$. Then there exists N such that $\phi \le f_n$ for all $n \ge N$. Hence,

$$\int \phi \, d\mu \le \mathbb{E}[f_n] \quad \text{for all } n \ge N.$$

Taking $\liminf$,

$$\int \phi \, d\mu \le \liminf_{n\to\infty} \mathbb{E}[f_n].$$

Taking the supremum over all such $\phi \le f$, we get

$$\mathbb{E}[f] \le \liminf_{n\to\infty} \mathbb{E}[f_n].$$

On the other hand, $f_n \le f$ implies $\mathbb{E}[f_n] \le \mathbb{E}[f]$, so

$$\limsup_{n\to\infty} \mathbb{E}[f_n] \le \mathbb{E}[f].$$

Combining the two inequalities we have,

$$\lim_{n\to\infty} \mathbb{E}[f_n] = \mathbb{E}[f]$$

as desired. □

Remark 7.2. It is crucial that the functions f_n are non-negative and increasing. If either condition is relaxed, the conclusion may fail.

7.2. **Fatou's Lemma.**

Theorem 7.3 (Fatou's Lemma)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of non-negative measurable functions. Then:*

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

Proof. Define $Y_k = \inf_{n\ge k} X_n$. Then $Y_k \le X_n$ for all $n \ge k$, and the sequence (Y_k) is increasing:

$$Y_k = \inf_{n\ge k} X_n \le \inf_{n\ge k+1} X_n = Y_{k+1}.$$

We have $\lim_{k\to\infty} Y_k = \liminf_{n\to\infty} X_n$, and since $Y_k \uparrow$, the Monotone Convergence Theorem 7.1 gives:

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

Since $Y_k \le X_n$ for all $n \ge k$, it follows that:

$$\mathbb{E}[Y_k] \le \inf_{n\ge k} \mathbb{E}[X_n].$$

Taking limits:

$$\lim_{k\to\infty} \mathbb{E}[Y_k] \le \lim_{k\to\infty} \inf_{n\ge k} \mathbb{E}[X_n] = \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Thus,

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

as desired. □

8. Upcrossings and Convergence

A central idea in establishing almost sure convergence of submartingales is the concept of *upcrossings*. The fundamental idea is that if a process fails to converge, it must oscillate infinitely often between two levels. Upcrossings provide a quantitative measure of such oscillations [Wil91, Doo48].

8.1. Definition of Upcrossings.

Definition 8.1 (Upcrossings)**.** Let $(X_n)_{n\ge 0}$ be a real-valued stochastic process and let $a < b$ be real numbers. The number of *completed upcrossings* of the interval $[a, b]$ by time n, denoted $U_n(a, b)$, is defined as the maximal integer k such that there exist increasing indices

$$0 \le \tau_1 < \sigma_1 < \tau_2 < \sigma_2 < \cdots < \tau_k < \sigma_k \le n,$$

with

$$X_{\tau_i} \le a \quad \text{and} \quad X_{\sigma_i} \ge b \quad \text{for all } i = 1, \dots, k.$$

Remark 8.2. Intuitively, an upcrossing of $[a, b]$ occurs when the process falls below the level a and subsequently rises above the level b. The count $U_n(a, b)$ measures how many such full oscillations between a and b occur up to time n.

The fundamental result that bounds the expected number of upcrossings of a submartingale is the following lemma due to Doob [Doo48].

8.2. Doob's Upcrossing Lemma.

Theorem 8.3 (Doob's Upcrossing Lemma)**.** *Let $(X_n)_{n\ge 0}$ be a real-valued submartingale adapted to a filtration $(\mathcal{F}_n)$, and let $a < b$. Let $U_n(a, b)$ denote the number of completed upcrossings of $[a, b]$ by time n. Then:*

$$(b - a)\,\mathbb{E}[U_n(a, b)] \le \mathbb{E}[(X_n - a)^+],$$

where $(x)^+ = \max(x, 0)$.

Proof. Define the process $Z_n = (X_n - a)^+$. By Example 6.2, (Z_n) is also a submartingale, since the positive part of a submartingale remains a submartingale.

Let τ_{2k-1} denote the time of the k-th entry below a, and let τ_{2k} be the next time the process reaches or exceeds b. We define the indicator variable I_i to capture whether time i falls within an upcrossing interval:

$$I_i = \begin{cases} 1 & \text{if } \tau_{2k-1} < i \leq \tau_{2k} \text{ for some } k, \\ 0 & \text{otherwise.} \end{cases}$$

That is, $I_i = 1$ precisely when the process is between entering below a and exiting at or above b, making an upcrossing of the interval $[a, b]$.

Note that during each upcrossing, the process must increase from at most a to at least b, so the corresponding increase in $Z_n = (X_n - a)^+$ is at least $b - a$. Hence, the total increase in Z_n over all completed upcrossings by time n satisfies:

$$(b-a) \cdot U_n(a,b) \leq \sum_{i=1}^{n} (Z_i - Z_{i-1}) I_i,$$

Taking expectations:

$$(b-a)\, \mathbb{E}[U_n(a,b)] \leq \sum_{i=1}^{n} \mathbb{E}[(Z_i - Z_{i-1}) I_i].$$

Now, observe that each I_i is $\mathcal{F}_{i-1}$-measurable, and we can apply the tower property of conditional expectation:

$$\begin{aligned} \mathbb{E}[(Z_i - Z_{i-1}) I_i] &= \mathbb{E}\left[I_i \cdot \mathbb{E}[Z_i - Z_{i-1} \mid \mathcal{F}_{i-1}]\right] \\ &= \mathbb{E}\left[I_i \cdot (\mathbb{E}[Z_i \mid \mathcal{F}_{i-1}] - Z_{i-1})\right]. \end{aligned}$$

Since (Z_n) is a submartingale, $\mathbb{E}[Z_i \mid \mathcal{F}_{i-1}] \geq Z_{i-1}$, so the inner difference is nonnegative. Also, $I_i \in [0,1]$, so dropping I_i can only increase the expectation:

$$\mathbb{E}[(Z_i - Z_{i-1}) I_i] \leq \mathbb{E}[\mathbb{E}[Z_i \mid \mathcal{F}_{i-1}] - Z_{i-1}] = \mathbb{E}[Z_i] - \mathbb{E}[Z_{i-1}].$$

Summing over $i = 1$ to n, we obtain a telescoping sum:

$$\sum_{i=1}^{n} \mathbb{E}[(Z_i - Z_{i-1}) I_i] \leq \mathbb{E}[Z_n] - \mathbb{E}[Z_0].$$

Since $Z_0 = (X_0 - a)^+ \geq 0$, this implies:

$$(b-a)\, \mathbb{E}[U_n(a,b)] \leq \mathbb{E}[Z_n] - \mathbb{E}[Z_0] \leq \mathbb{E}[(X_n - a)^+].$$

This completes the proof. □

8.3. **Convergence from Finite Upcrossings.** We now connect upcrossings to almost sure convergence.

Theorem 8.4 (Upcrossing Criterion for Convergence)**.** *Let $(X_n)_{n\geq 0}$ be a real-valued submartingale adapted to a filtration $(\mathcal{F}_n)$. Suppose that for every rational pair $a < b$,*

$$\mathbb{P}(U(a,b) < \infty) = 1,$$

where $U(a,b)$ is the total number of upcrossings of $[a,b]$. Then the sequence (X_n) converges almost surely to a finite random variable:

$$\mathbb{P}\left(\lim_{n\to\infty} X_n \text{ exists and is finite}\right) = 1.$$

Proof. Suppose the limit $\lim_{n\to\infty} X_n$ does not exist on a set of positive probability. Then there exists $\omega \in \Omega$ such that:

$$\liminf_{n\to\infty} X_n(\omega) < \limsup_{n\to\infty} X_n(\omega).$$

As the rational numbers are dense over $\mathbb{R}$, we can choose rationals $a < b$ such that:

$$\liminf X_n(\omega) < a < b < \limsup X_n(\omega).$$

Then the trajectory $X_n(\omega)$ must cross from below a to above b infinitely often, implying $U(a,b)(\omega) = \infty$, contradicting the assumption that $\mathbb{P}(U(a,b) < \infty) = 1$. Hence, with probability one, (X_n) converges to a finite limit X_∞. Since submartingales are integrable and $X_n \to X_\infty$ a.s., we conclude that $X_\infty < \infty$ almost surely. □

9. Martingale Convergence Theorem

The Martingale Convergence Theorem is a cornerstone result in probability theory. It guarantees that martingales and submartingales with uniformly bounded expectations converge almost surely and in L^1 [Doo48, Dur19, Wil91]. This theorem is a culmination of the previous results involving upcrossings and measure-theoretic convergence theorems.

Theorem 9.1 (Martingale Convergence Theorem). *Let $(X_n)_{n\geq 0}$ be a real-valued submartingale adapted to a filtration $(\mathcal{F}_n)$, and suppose*

$$\sup_n \mathbb{E}[|X_n|] = M < \infty.$$

Then:

1. *(X_n) converges almost surely to a finite random variable X_∞,*
2. *$X_\infty \in L^1$, i.e., $\mathbb{E}[|X_\infty|] < \infty$.*

Proof. Fix $a < b$ and define $U_n(a,b)$ as the number of completed upcrossings of $[a,b]$ by time n. By Doob's Upcrossing Lemma 8.3:

$$(b-a)\cdot \mathbb{E}[U_n(a,b)] \leq \mathbb{E}[(X_n - a)^+].$$

Note that $(X_n - a)^+ \leq |X_n| + |a|$, so:

$$\mathbb{E}[U_n(a,b)] \leq \frac{\mathbb{E}[|X_n|] + |a|}{b-a} \leq \frac{M + |a|}{b-a}.$$

Taking $n \to \infty$, we conclude:

$$\mathbb{E}[U(a,b)] = \lim_{n\to\infty} \mathbb{E}[U_n(a,b)] \leq \frac{M+|a|}{b-a} < \infty,$$

so $U(a,b) < \infty$ almost surely. Since this holds for all rational $a < b$, Theorem 8.4 implies that (X_n) converges almost surely to some finite limit X_∞.

To show $X_\infty \in L^1$, apply Fatou's Lemma 7.2:

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

Thus, X_∞ is integrable. □

Remark 9.2. The L^1-boundedness condition is essential and cannot be dropped. For instance, let $S_n = \sum_{k=1}^{n} Y_k$ be a symmetric random walk where $Y_k \in \{-1, 1\}$ are i.i.d. with mean zero. Then (S_n) is a martingale, but $\mathbb{E}[|S_n|] \sim \sqrt{n} \to \infty$. The process does not converge almost surely, demonstrating that boundedness in expectation is a necessary condition for convergence.

Example (Convergence of a Bounded Martingale). Let (X_n) be a simple symmetric random walk on $\mathbb{Z}$ starting at 0, and let τ be the hitting time of either -5 or $+5$. Define the stopped process $Y_n = X_{n\wedge\tau}$. Then (Y_n) is a martingale bounded between -5 and 5, so

$$\sup_n \mathbb{E}[|Y_n|] \leq 5.$$

By the Martingale Convergence Theorem, $Y_n \to Y_\infty$ a.s. and in L^1. Indeed, Y_n converges almost surely to the value $Y_\infty = \pm 5$ depending on whether the walk hits $+5$ or -5 first.

We now present a useful corollary, which simplifies the conditions when the martingale is either bounded or non-negative.

Corollary 9.3 (Convergence of Bounded or Non-negative Martingales). *Let $(X_n)_{n\geq 0}$ be a real-valued martingale adapted to a filtration $(\mathcal{F}_n)$. Suppose either:*

- *$X_n \geq C$ for all n, or*
- *$X_n \leq C$ for all n,*

for some constant $C \in \mathbb{R}$. Then there exists a finite random variable X_∞ such that:

$$X_n \xrightarrow{a.s.} X_\infty.$$

Proof. First, consider the case when $X_n \geq 0$ for all n. Then, the expectations $\mathbb{E}[X_n]$ are constant:

$$\mathbb{E}[|X_n|] = \mathbb{E}[X_n] = \mathbb{E}[X_0], \quad \text{for all } n,$$

so (X_n) is uniformly bounded in L^1. Hence, by the Martingale Convergence Theorem 9.1, $X_n \to X_\infty$ almost surely for some finite X_∞.

Now suppose $X_n \geq C$ for some $C \in \mathbb{R}$. Define the shifted process $Y_n = X_n - C$, which satisfies $Y_n \geq 0$. Since (X_n) is a martingale, so is (Y_n), and by the same argument, $Y_n \to Y_\infty$ a.s. for some finite Y_∞. Hence,

$$X_n = Y_n + C \to Y_\infty + C = X_\infty \quad \text{a.s.}$$

Similarly, if $X_n \leq C$, define $Z_n = C - X_n \geq 0$, so that (Z_n) is a non-negative martingale. Then $Z_n \to Z_\infty$ a.s., and

$$X_n = C - Z_n \to C - Z_\infty = X_\infty \quad \text{a.s.}$$

as desired. □

Remark 9.4. This corollary is frequently used in applications, where martingales arise with natural lower or upper bounds—for example, likelihood ratios in hypothesis testing or capital in gambling systems.

Example (Martingale with Non-negative Increments: Gambling Game). Let a gambler's wealth process be modeled by a martingale (X_n), where the player starts with \$1 and repeatedly bets all their wealth on a fair coin flip. If they win, they double their money; if they lose, they go bankrupt. The wealth process can be described as:

$$X_{n+1} = \begin{cases} 2X_n & \text{with probability } 1/2, \\ 0 & \text{with probability } 1/2, \end{cases} \quad \text{with } X_0 = 1.$$

This process is not bounded in L^1, since:

$$\mathbb{E}[X_n] = 1 \quad \text{for all } n,$$

but

$$\mathbb{E}[|X_n|] = 1 \quad \text{and} \quad \sup_n \mathbb{E}[|X_n|] = 1.$$

So it satisfies the Martingale Convergence Theorem. In fact, $X_n \to X_\infty = 0$ almost surely, since the probability of bankruptcy tends to 1 as $n \to \infty$.

Example. Let $(Z_n)_{n\geq 0}$ be a Galton–Watson branching process with $Z_0 = 1$, and let $m = \mathbb{E}[X]$ be the expected number of offspring for a single individual, where X is the offspring distribution. Assume $m < \infty$. Define the normalized process

$$M_n = \frac{Z_n}{m^n}.$$

Then, $(M_n)_{n\geq 0}$ is a non-negative martingale with respect to the natural filtration $(\mathcal{F}_n)$. By the Martingale Convergence Theorem, since $M_n \geq 0$ and $\mathbb{E}[M_n] = 1$, we have:

$$M_n \xrightarrow{a.s.} M_\infty \quad \text{as } n \to \infty,$$

for some random variable M_∞. Moreover, M_∞ is $\mathcal{F}_\infty$-measurable and satisfies $\mathbb{E}[M_\infty] \leq 1$.

Example (Failure Without L^1-Boundedness)*.* Define $X_n = n \cdot \mathbf{1}_{\{T\leq n\}}$, where T is a random variable with tail:

$$\mathbb{P}(T > n) = \frac{1}{\log(n+2)}.$$

Then (X_n) is adapted and increasing, hence a submartingale. However,

$$\mathbb{E}[X_n] = n \cdot \mathbb{P}(T \leq n) = n\left(1 - \frac{1}{\log(n+2)}\right) \to \infty,$$

so $\sup_n \mathbb{E}[|X_n|] = \infty$, and the Martingale Convergence Theorem does not apply. In fact, $X_n \to \infty$ almost surely, and does not converge.

10. Extensions and Applications

10.1. **Doob's L^p Convergence Theorem.** The martingale convergence theorem guarantees almost sure and L^1 convergence under boundedness in expectation. Doob's L^p version strengthens this by requiring boundedness in higher moments, and concludes convergence in the stronger L^p-norm [Doo53, Dur19].

Theorem 10.1 (Doob's L^p Convergence Theorem)**.** *Let $(X_n)_{n\geq 0}$ be a real-valued submartingale (or martingale) adapted to a filtration $(\mathcal{F}_n)$. Suppose $1 < p < \infty$ and*

$$\sup_n \mathbb{E}[|X_n|^p] < \infty.$$

Then:

(1) *$X_n \to X_\infty$ almost surely,*
(2) *$X_n \to X_\infty$ in L^p, i.e., $\|X_n - X_\infty\|_p \to 0$,*

(3) *The limit satisfies* $\mathbb{E}[|X_\infty|^p] < \infty$.

Remark 10.2. This result strengthens the Martingale Convergence Theorem by providing convergence in L^p, not just in probability or almost surely. It is particularly useful when working with square-integrable martingales and appears frequently in stochastic calculus.

Example. Let $(B_t)_{t\geq 0}$ be standard Brownian motion and fix $\lambda \in \mathbb{R}$. Define the exponential martingale

$$M_t = \exp\left(\lambda B_t - \frac{\lambda^2}{2}t\right).$$

Then $(M_t)_{t\geq 0}$ is a positive martingale. For any $p > 1$, we can compute:

$$\mathbb{E}[M_t^p] = \mathbb{E}\left[\exp\left(p\lambda B_t - \frac{p\lambda^2}{2}t\right)\right] = \exp\left(\frac{p^2\lambda^2 t}{2} - \frac{p\lambda^2 t}{2}\right) = \exp\left(\frac{p(p-1)\lambda^2 t}{2}\right),$$

which is finite only if t is sufficiently small. However, for small time horizons t, M_t is bounded in L^p, so:

$$M_t \xrightarrow{L^p} M_\infty \quad \text{as } t \to \infty.$$

10.2. **Continuous-Time Extension.** The Martingale Convergence Theorem extends naturally to continuous-time processes, provided we work in an appropriate framework [KS98]. Let $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions (i.e., the filtration is right-continuous and complete).

Definition 10.3 (Continuous-Time Martingale)**.** A stochastic process $(X_t)_{t\geq 0}$ is a *continuous-time martingale* with respect to the filtration $(\mathcal{F}_t)$ if:

(1) X_t is $\mathcal{F}_t$-measurable for each $t \geq 0$ (i.e., the process is adapted),
(2) $\mathbb{E}[|X_t|] < \infty$ for all $t \geq 0$,
(3) For all $0 \leq s \leq t$, we have

$$\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s \quad \text{almost surely.}$$

If the inequality $\mathbb{E}[X_t \mid \mathcal{F}_s] \geq X_s$ holds instead, the process is a *submartingale*; if the inequality is reversed, it is a *supermartingale*.

Theorem 10.4 (Continuous-Time Martingale Convergence)**.** *Let $(X_t)_{t\geq 0}$ be a real-valued, right-continuous submartingale adapted to $(\mathcal{F}_t)$, and suppose:*

$$\sup_{t\geq 0} \mathbb{E}[|X_t|] < \infty.$$

Then there exists an integrable random variable X_∞ such that:

$$X_t \to X_\infty \quad \textit{almost surely as } t \to \infty.$$

Moreover, if (X_t) is a uniformly integrable martingale, then convergence also holds in L^1.

Example (Brownian Motion Stopped at an Integrable Time). Let $(B_t)_{t\geq 0}$ be standard Brownian motion. For any stopping time τ with $\mathbb{E}[\tau] < \infty$, the stopped process $(B_{t\wedge\tau})_{t\geq 0}$ is a uniformly integrable martingale. By the Martingale Convergence Theorem in continuous time, we have:

$$B_{t\wedge\tau} \to B_\tau \quad \text{almost surely and in } L^1 \text{ as } t \to \infty.$$

This convergence underpins many classical results in stochastic calculus, including the optional stopping theorem and properties of predictable processes.

10.3. **Application: Martingales in Mathematical Finance.** Martingales lie at the heart of modern financial mathematics. Under a *risk-neutral measure*, the prices of tradable assets—when properly discounted—evolve as martingales. This fundamental principle ensures that the market is free of arbitrage opportunities and allows for the fair pricing of financial derivatives [EK05].

Example (Pricing a European Call Option via Martingale Representation)*.* Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{Q})$ be a filtered probability space satisfying the usual conditions, where $\mathbb{Q}$ denotes the risk-neutral probability measure. Let $(S_t)_{t\geq 0}$ be the price process of a stock, modeled by the stochastic differential equation:

$$dS_t = rS_t\,dt + \sigma S_t\,dW_t,$$

under the measure $\mathbb{Q}$, where:

- r is the constant risk-free interest rate,
- $\sigma > 0$ is the volatility of the stock,
- W_t is a Brownian motion under $\mathbb{Q}$.

This model is known as the Black-Scholes model. Now let $B_t = e^{rt}$ represent the value of a risk-free bond (bank account), which grows deterministically at rate r. We define the *discounted stock price* by:

$$\tilde{S}_t = \frac{S_t}{B_t} = S_t e^{-rt}.$$

Applying Itô's lemma, we can show that the drift term in $\tilde{S}_t$ vanishes, and thus $(\tilde{S}_t)_{t\geq 0}$ is a $\mathbb{Q}$-martingale. This reflects the key financial insight: *under the risk-neutral measure, the expected future discounted price of an asset is its current price.*

Now consider a European call option with strike price $K > 0$ and maturity $T > 0$. Its payoff at maturity is given by:

$$H = \max(S_T - K, 0),$$

which is the amount the option holder gains if the stock price exceeds the strike price at time T, and zero otherwise.

By the fundamental theorem of asset pricing, the arbitrage-free price of this option at an earlier time $t \leq T$ is the expected discounted payoff under $\mathbb{Q}$, conditional on current information $\mathcal{F}_t$:

$$V_t = \mathbb{E}^{\mathbb{Q}}\left[\frac{H}{B_T}\,\middle|\,\mathcal{F}_t\right].$$

This defines a stochastic process $(V_t)_{t\in[0,T]}$, which represents the option's price through time. Since conditional expectation of a square-integrable (or integrable) random variable defines a martingale, the process (V_t) is itself a $\mathbb{Q}$-martingale.

The Martingale Convergence Theorem now tells us that because (V_t) is a non-negative L^1-bounded martingale, it must converge almost surely and in $L^1(\mathbb{Q})$ to:

$$\lim_{t\to T} V_t = V_T = \frac{H}{B_T}.$$

That is, the price process (V_t) evolves in a way that ensures it converges to the actual payoff (appropriately discounted) at maturity. This formalizes the idea that the option price "tracks" the future payoff as more information is revealed over time.

From a financial perspective, this convergence ensures that as we approach maturity, the option's value increasingly reflects the known final outcome. If the stock ends above the strike, the option's value approaches the in-the-money amount; otherwise, it converges to zero. This behavior is not just expected — it is guaranteed by the Martingale Convergence Theorem under the no-arbitrage pricing framework.

10.4. **Application: Martingale Proof of the Strong Law of Large Numbers.** The Martingale Convergence Theorem can be used to give an elegant proof of the *Strong Law of Large Numbers* (SLLN) under mild moment conditions, illustrating its strength beyond direct modeling applications [Dur19].

Theorem 10.5 (Strong Law via Martingales)**.** *Let* $(X_n)_{n\geq 1}$ *be a sequence of i.i.d. random variables with* $\mathbb{E}[X_1] = \mu$ *and* $\mathrm{Var}(X_1) < \infty$*. Then*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} \mu.$$

Proof. We define a new process by first centering the variables:

$$Y_k = X_k - \mu,$$

so that $\mathbb{E}[Y_k] = 0$. Define the filtration $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ and the sequence:

$$M_n = \sum_{k=1}^{n} Y_k.$$

This is a martingale with respect to $(\mathcal{F}_n)$, since:

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = M_n + \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] = M_n.$$

We now normalize:

$$A_n = \frac{M_n}{n} = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu) = \frac{1}{n}\sum_{k=1}^{n} X_k - \mu.$$

To prove $A_n \to 0$ almost surely (which implies the SLLN), we can apply the Martingale Convergence Theorem 9.1 directly. Define the stopped martingale:

$$M_n^{(T)} = M_{\min(n,T)},$$

where T is a fixed index. Since M_n has bounded second moment, the MCT guarantees that:

$$M_n \xrightarrow{a.s.} M_\infty \quad \text{and} \quad \frac{M_n}{n} \xrightarrow{a.s.} 0.$$

Thus,

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s.} \mu,$$

as desired. □

Remark 10.6. This martingale-based proof highlights how viewing partial sums of zero-mean i.i.d. variables as martingales allows the elegant use of convergence theorems. While traditional proofs invoke *Kolmogorov's Three Series Theorem* or *Chebyshev's Inequality* with the *Borel–Cantelli Lemma*, the martingale route provides a unifying probabilistic perspective.

11. Conclusion

The Martingale Convergence Theorem stands as a cornerstone of modern probability theory, revealing the deep regularity underlying fair or unfavorable stochastic processes under minimal boundedness assumptions. In this paper, we have:

- Developed the measure-theoretic framework for probability, including filtrations, conditional expectations, and integrability,
- Introduced martingales, submartingales, and supermartingales through intuitive examples such as random walks and conditional projections,
- Proved key analytic tools, including Jensen's inequality, Fatou's Lemma, and the Monotone Convergence Theorem,
- Established control over pathwise oscillations via Doob's Upcrossing Lemma,
- Derived the Martingale Convergence Theorem and its corollaries for bounded and non-negative martingales,
- Explored extensions to continuous time and convergence in L^p, as well as applications to financial mathematics.

References

[Axl22] Sheldon Axler. *Measure Theory: A First Course*. Springer, Cham, 2nd edition, 2022.
[Bil95] Patrick Billingsley. *Probability and Measure*. Wiley, New York, 3rd edition, 1995.
[Doo48] Joseph L. Doob. The martingale convergence theorem. *Trans. Amer. Math. Soc.*, 63(3):373–386, 1948.
[Doo53] Joseph L. Doob. *Stochastic Processes*. Wiley, 1953.
[Dur19] Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 5th edition, 2019.
[EK05] Robert J. Elliott and P. Ekkehard Kopp. *Mathematics of Financial Markets*. Springer, New York, 2nd edition, 2005.
[Kle13] Achim Klenke. *Probability Theory: A Comprehensive Course*. Springer, 2nd edition, 2013.
[Kol56] Andrey Nikolaevich Kolmogorov. *Foundations of the Theory of Probability*. Chelsea Publishing Company, New York, 1956.
[KS98] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 2nd edition, 1998.
[RF10] Halsey Royden and Patrick Fitzpatrick. *Real Analysis*. Pearson, Boston, 4th edition, 2010.
[Wil91] David Williams. *Probability with Martingales*. Cambridge University Press, 1991.