

MARTINGALES

ROHAN SHAH

Abstract. This paper covers martingales with introduction to measure theory concepts, and other concepts, including the Lebesgue Integration and Conditional Expectation. It follows up with proofs on Kolmogorov's Theorem on conditional expectations, the Martingale Property, and the Pythagorean Theorem on Martingales. Finally, it ends with martingales' application in finance.

CONTENTS

1. INTRODUCTION	1
2. PROBABILITY AND MEASURE THEORY	2
3. LEBESGUE INTEGRATION	5
4. CONDITIONAL EXPECTATION	7
5. MARTINGALES	11
6. APPLICATIONS IN FINANCE	17
7. ACKNOWLEDGEMENTS	17
References	18

1. INTRODUCTION

Let's consider a bettor participating in a game involving the flipping of a fair coin. In this game, the bettor earns \$1 for a heads outcome and loses \$1 for a tails outcome. This means that if the better flips the coin once and it lands on

Date: July 16, 2024.

tails, they would lose a dollar. Conversely, if it lands on heads, they would gain a dollar.

Over many run-throughs of flipping the coin, the bettor's total earnings, represented as T , could be either positive or negative, depending on the outcomes of the flips. Each individual coin flip carries an equal probability of 50%, meaning there is an equal chance that the next flip will result in either heads or tails. Consequently, with each flip, the bettor's total earnings will either increase by \$1 (if heads) or decrease by \$1 (if tails).

After many flips, the bettor's total earnings will result from random fluctuations due to the equal probability of heads or tails. Over many flips, the total earnings T will show a random pattern of gains and losses, yet the expectation of total earnings remains zero. This means that at one point, despite the variances in time, basic probability states that the final value will be 0. This is an example of a fair game, where the expected earnings after the next turn would equal the current earnings. A martingale shares a similar idea. Introduced by Paul Lévy in 1934 and later by Ville in 1939, a martingale is a probability model based on a fair game. This property can be translated to concepts outside regular games, including the stock market. In simple terms, a martingale is a process where the conditional expectation of the next value is equal to the present value. Since there are no predictable trends in a fair game, a martingale can effectively visualize the randomness and fairness of such games. This paper covers the mathematical definition of martingale with the requirement of knowledge of elementary probability, specifically measure theory (σ -Algebra, measurable spaces, Lebesgue measure, Borel sets).

Before we begin, I will introduce the notation for a martingale and explain each part throughout the paper.

Definition 1.1. X is a martingale if

$$X_n = E[X_{n+1} | \mathcal{F}_n]$$

2. PROBABILITY AND MEASURE THEORY

We begin with definitions regarding the basics of measure theory, including sets and events. We are given a sample space denoted by Ω as the set of all possible outcomes of a random experiment. An event is a subset of the sample space Ω .

For instance, if Ω represents the possible outcomes of a dice roll, then $(2,4,6)$ could be an event representing the roll of an even number. We begin by looking at the first approximation, the power set of omega, $P(\Omega)$ which equals the set of all subsets of Ω . To restrict this power set, we introduce our first definition.

Definition 2.1. A system $A \subseteq P(X)$, where A is a collection of elements of subsets of X. This collection is called σ -algebra. A σ -algebra is a collection of subsets of the sample space that includes the empty set and the sample space itself. However, to be named a σ -algebra (an important measure theory definition), it must follow these three rules:

- $\emptyset, X \subseteq P(X)$
- If $B \in A$ then $B^C = X/B \in A$. This means that for any set in the σ -algebra, its complement is also in the σ -algebra.
- $B, D \in A \rightarrow B \cup D \in A$. Additionally, if the set B and D are in set, A, the intersection of set B and D will also be in A.
- Then (X, A) is a measurable space.

Measurable spaces are essential to provide a framework in which measures are defined. In the definition above, (X, A) is a measurable space with set X and the collection of subsets, the sigma-algebra, A. The sigma-algebra includes the subsets of X, that are considered measurable.

Definition 2.2.

- For $M \subseteq P(X)$, there is a smallest sigma algebra that contains M:

$$\sigma(M) =: \bigcap_{A \supseteq M} A \text{ is a } \sigma\text{-algebra generated by set M.}$$

- E.g. $X = \{a, b, c, d\}, M = \{\{a\}, \{b\}\}$
- Therefore, $\sigma(\mathcal{M}) = \left\{ \phi, X, \{a\}, \{b\}, \{a, b\}, \{b, c, d\}, \{a, c, d\}, \{c, d\} \right\}$

By defining $\sigma(M)$ as the intersection of all sigma-algebras containing M , we guarantee that $(\sigma(M))$ includes all necessary sets to be a sigma-algebra while being minimal in the sense that it does not include any extraneous sets not required by the sigma-algebra properties and the inclusion of M .

Definition 2.3. Let A be an algebra, $\mu : A \rightarrow [0, \infty]$ is a finitely additive measure on A, if

- For any two sets $B, D \in A$ with $B \cap D = \emptyset$: $\mu(B \cup D) = \mu(B) + \mu(D)$.

A finitely additive measure is a function that assigns a non-negative value to each set, representing its size, in a way that the measure of the union of two non-overlapping sets equals the sum of their measures. This property ensures consistent and additive measurements, allowing us to determine the size of complex shapes by summing the measures of their non-overlapping parts.

Proof. Using Definition 2.1 and 2.3:

- Then $A_1, A_2, A_3, \dots, A_n \in A$: $A_1 \cup A_2 \cup \dots \cup A_n \in A$
- If $A_i \cap A_j = \emptyset$ $1 < i \neq j < n$, then $\mu \bigcup_{i=1}^n A_i = \sum_{i=1}^n \mu(A_i)$.

The proof shows that any finite union of sets in a sigma-algebra A is also in \mathcal{A} . It then uses the countable additivity property of measures to show that the measure of a finite union of disjoint sets is the sum of the measures of the individual sets.

Definition 2.4. A measure μ on an algebra, A is a set function from $A \rightarrow [0, \infty]$ if it follows the following properties:

- If we take any sequence of sets: $A_1, A_2, A_3, \dots, A_n \in A$ such that $\bigcup_{i=1}^n A_i$ and $A_i \cap A_j = \emptyset$ if $i \neq j$ then $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$.

Note that this property is denoted as σ -additivity. However if the set function, A is in the interval $[0, 1]$, then these properties won't apply as it is a probability measure rather than a measure. Now, we can define what a probability measure is.

Definition 2.5. A probability measure (P) on the sample space satisfies $P(\Omega) = 1$. Therefore, the triple (Ω, \mathcal{F}, P) is called a probability space, where \mathcal{F} is the sigma algebra and P is the probability measure.

With these terms defined, we can introduce what random variables are.

Definition 2.6. Given a measurable space (Ω, \mathcal{F}) , a function $X : \Omega \rightarrow \mathbb{R}$ is said to be a measurable function or a random variable if for every Borel Set, $B \in \mathbb{B}(\mathbb{R})$, the pre-image of that set, $X^{-1}(B) \in \mathcal{F}$. This implies that the function

X maps events in the sample space Ω to real numbers in a way that preserves the structure of σ -algebra.

Definition 2.7. A random variable is a measurable function from the sample space to the real numbers. The distribution of a random variable X is a function $X : B \rightarrow [0, 1]$ such that $X(B) = P(X^{-1}(B))$ for all Borel sets B . Note that Borel sets (subset) are important as a way of showing that if open sets are measurable, then so are all other sets of interest (like closed sets, countable unions, and intersections).

3. LEBESGUE INTEGRATION

If X , a integrable random variable takes on an infinite number of values, calculating the weighted average requires integration to handle this infinite sum. However, the standard Riemann integral is often unsuitable in this context. The Riemann integral approximates the integral by splitting the domain into contiguous intervals to form "rectangles," then summing the areas of these rectangles. This approach might not make sense for an abstract sample space, and some measurable functions may not meet the continuity requirements for Riemann integration. Therefore, we use a more generalized form of the Riemann integral called the Lebesgue integral. However, before we define the integral, we need understand simple functions.

Definition 3.1. Let A be a subset of Ω . The indicator function of A is the function $\mathbf{1}_A : \Omega \rightarrow \mathbb{R}$ defined as:

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}.$$

Theorem 3.2. *The indicator function of A is measurable in every σ -algebra \mathcal{F} containing A .*

Proof. Every possible Borel set B falls into one of four categories:

- $1 \in B, 0 \notin B$. Then $\mathbf{1}_A^{-1}(B) = A$;
- $0 \in B, 1 \notin B$. Then $\mathbf{1}_A^{-1}(B) = A^c$;
- $0, 1 \in B$. Then $\mathbf{1}_A^{-1}(B) = \Omega$;
- $0, 1 \notin B$. Then $\mathbf{1}_A^{-1}(B) = \emptyset$.

By definition, if $A \in \mathcal{F}$, then $\Omega, \emptyset, A^c \in \mathcal{F}$. Thus, by Definition 3.1, $\mathbf{1}_A$ is measurable.

Definition 3.3. Simple functions are measurable functions that take on a finite number of values. They can be written as:

$$g = \sum_{i=1}^n a_i \mathcal{X}_{A_i}$$

In this notation above, a_i are real numbers, A_i are measurable sets, and \mathcal{X}_{A_i} .

Proof. If we integrate the following simple function using the linearity property:

1. Define the Simple Function:

$$g = \sum_{i=1}^n a_i \mathcal{X}_{A_i}$$

2. Set Up the Integral: To integrate g with respect to a measure μ , we use the definition:

$$\int g d\mu = \int \sum_{i=1}^n a_i \mathcal{X}_{A_i} d\mu$$

3. Use Linearity of the Integral: By the linearity property of the integral, we can move the summation outside the integral:

$$\int \sum_{i=1}^n a_i \mathcal{X}_{A_i} d\mu = \sum_{i=1}^n a_i \int \mathcal{X}_{A_i} d\mu$$

4. Evaluate Each Integral: The integral of the characteristic function \mathcal{X}_{A_i} of a set A_i with respect to μ is simply the measure of A_i :

$$\int \mathcal{X}_{A_i} d\mu = \mu(A_i)$$

5. Combine the Results: Substitute the measure of each A_i into the sum:

$$\int g d\mu = \sum_{i=1}^n a_i \mu(A_i)$$

So, the integral of the simple function g with respect to the measure μ is:

$$\int g d\mu = \sum_{i=1}^n a_i \mu(A_i), \text{ where } \mu \text{ is a measure on } \Omega$$

Properties of Lebesgue Integral:

- Linearity: $\int (af + bg)d\mu = a \int fd\mu + b \int gd\mu$
- Monotonicity: If $f \leq g$, then $\int fd\mu \leq \int gd\mu$
- Countable Additivity: $\int \sum_{i=1}^{\infty} f_i d\mu = \sum_{i=1}^{\infty} \int f_i d\mu$

The utility of the Lebesgue integral becomes clearer with this definition. The Lebesgue integral of a simple function is simply the sum of its possible values weighted by the measure of the subsets of the domain corresponding to those values. In the context of a probability measure, this is equivalent to averaging the possible values of a random variable weighted by the probability of the events that produce those values—that is, the expectation. For more complex, non-simple functions, the integral can be defined by approximating the function with simple functions. This leads to the following theorem, which facilitates this process.

Theorem 3.4. *Let $f : \Omega \rightarrow [0, \infty]$ be a measurable function on the space $(\Omega, \mathcal{F}, \mu)$. Then, there exists an increasing sequence of simple functions $(x_i)_{i=1}^{\infty}$ that converges point-wise to f .*

This is known as the Simple Function Approximation Theorem. As we covered a little about expectation, we can now understand both that and conditional expectation.

4. CONDITIONAL EXPECTATION

Conditional expectation, denoted $E[Y|X]$, represents the expected value of a random variable Y given the knowledge of another variable X . It is used to update our expectation of Y based on the additional information provided by X . To understand this better, we begin by defining simple random variables, then move to expectation.

Definition 4.1. A random variable, X is simple if there exists $n > 0, X_1, X_2, \dots, X_n \in \mathbb{R}, A_1, A_2, \dots, A_n \in \mathcal{F}$:

$$X = \sum_{X_i=1}^n X_i \mathbb{1}_{A_i}$$

This means that X is a function from $\Omega \rightarrow \mathbb{R}$. A_i are sets in the σ -algebra, meaning that they are subsets in Ω .

Theorem 4.2. For every random variable X , there exist X_1, X_2, \dots that are known as simple random variables such that:

$$|X_n| \leq |X|$$

For every $\omega \in$ sample space, Ω .

$$X_n(\omega) \nearrow_{\infty} X(\omega)$$

Theorem 4.3. If $X(\omega) \geq 0$ for all $\omega \in \Omega$, then $X_n(\omega)$ can be chosen to be non-decreasing in n for every ω .

”Non-decreasing for n ” means that as the sequence progresses, the values do not decrease for any outcome in the sample space. In other words, the values either stay the same or increase as the index increases, ensuring the sequence never gets smaller for each outcome.

After these definitions of random variables, we can now define the expectation of a random variable.

Definition 4.4. If $X = \sum_{X_i=1}^n X_i \mathbb{1}_{A_\varepsilon}$ is simple, then $\mathbb{E}X$ (expected value) = $\sum_{X_i=1}^n X_i \mathbb{P}(A_\varepsilon)$

This is Step 1, only applying if the random variable, X is a simple, following the properties defined in Definition 3.1. Now, we will discuss Step 2 when $X \geq 0$ (not negative).

Definition 4.5. If $X \geq 0$ and a random variable, then X_1, X_2, \dots is a simple random variable such that X_i converges to X .

$$\mathbb{E}X = \lim_{X \rightarrow i} \mathbb{E}X_i \in [0, \infty]$$

Definition 4.6. If X is a random variable, then $X := X^+ - X^-$

This second part of the definition is stating that the positive part of the variable, X^+ plus the negative part of the variable, X^- equals to the variable itself. For instance the positive part of 4 is 4, but the negative part of 4 is 0. Adding these two numbers, we get the number 4 itself. This may seem intuitive, but if we do this for every ω , we can define a function, where both the positive and negative

part of the variable is not negative. If we take the expectation of this equation we get the following:

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

unless

$$\mathbb{E}X^+ = \infty = \mathbb{E}X^-$$

Making $\mathbb{E}[X]$ not exist, which is known as the Cauchy random variable.

Now we move to defining what conditional expectation is.

Theorem 4.7. *Kolmogorov's Theorem on conditional expectations.*

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X is an integrable random variable; $\mathbb{E}|X| < \infty$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub sigma-algebra. Take random variable Y such that it follows these properties:

- Y is \mathcal{G} -measurable.
- $\mathbb{E}|Y| < \infty$
- $\forall A \in \mathcal{G}$
- $\int_A X dP = \int_A Y dP = \int_A \mathbb{E}(X|\mathcal{G}) dP$

The random variable Y is denoted by $\mathbb{E}(X|\mathcal{G})$ and is called the conditional expectation of X given \mathcal{G} .

The best way to explain this is through a proof, where we shall require the use of Radon-Nikodym Theorem.

Proof. Define a new measure, Q on (Ω, \mathcal{G}) by:

$$Q(A) = \int_A X dP$$

for all $A \in \mathcal{G}$.

Since X is integrable, Q is a finite measure on (Ω, \mathcal{G}) :

$$Q(\Omega) = \int_{\Omega} X dP = \mathbb{E}[X] < \infty$$

The Radon-Nikodym theorem states that since Q is a sigma-finite measure on

(Ω, \mathcal{G}) and $P|_{\mathcal{G}}$ is the restriction of the probability measure P to \mathcal{G} , there exists a \mathcal{G} -measurable function Y such that:

$$Q(A) = \int_A Y dP \quad \text{for all } A \in \mathcal{G}$$

From here, since $A \in \mathcal{G}$:

$$\int_A X dP = \int_A Y dP$$

Define $Y = \mathbb{E}[X | \mathcal{G}]$. By construction, Y is \mathcal{G} -measurable and satisfies:

$$\int_A X dP = \int_A \mathbb{E}[X | \mathcal{G}] dP \quad \text{for all } A \in \mathcal{G}$$

To show uniqueness, suppose Y' is another \mathcal{G} -measurable function that satisfies the same integral property. Then for all $A \in \mathcal{G}$:

$$\int_A \mathbb{E}[X | \mathcal{G}] dP = \int_A Y' dP$$

Taking $A = \{\mathbb{E}[X | \mathcal{G}] > Y'\}$, we get:

$$\int_{\{\mathbb{E}[X | \mathcal{G}] > Y'\}} (\mathbb{E}[X | \mathcal{G}] - Y') dP = 0$$

Since $\mathbb{E}[X | \mathcal{G}] - Y' \geq 0$ on $\{\mathbb{E}[X | \mathcal{G}] > Y'\}$, this implies $P(\{\mathbb{E}[X | \mathcal{G}] > Y'\}) = 0$. A similar argument holds for $\{\mathbb{E}[X | \mathcal{G}] < Y'\}$. Thus, $\mathbb{E}[X | \mathcal{G}] = Y'$ almost surely.

The Radon-Nikodym theorem provides a \mathcal{G} -measurable function Y that satisfies the integral property, ensuring the existence of the conditional expectation $\mathbb{E}[X | \mathcal{G}]$. The uniqueness follows from the fact that any two \mathcal{G} -measurable functions satisfying this property must be equal almost surely. Thus, Kolmogorov's theorem on conditional expectations is proved.

Let's move on to an important property of Conditional Expectation, the tower rule.

Definition 4.8. $(\Omega, \mathcal{F}, \mathbb{P})$; $\mathcal{G} \subseteq \mathcal{H} \subseteq \mathcal{F}$ (These are all σ -algebras). Both \mathcal{G} and \mathcal{H} are sub-sigma-algebras. Let X be a random variable, $\mathbb{E}[X] < \infty$.

Then the Tower rule states:

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}[X|\mathcal{G}]$$

The Tower Rule essentially tells us that if we first condition on a larger sigma-algebra \mathcal{G} and then on a smaller sigma-algebra \mathcal{H} , the result is the same as if we had conditioned on the smaller sigma-algebra \mathcal{H} directly.

5. MARTINGALES

We are given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration as defined below.

Definition 5.1. A filtration $(\mathcal{F}_t)_{t \geq 0}$ is an increasing family of σ -algebras:

$$\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F} \text{ for } s \leq t$$

From this we can define adaptation.

Definition 5.2. A sequence of random variables $\{X_n; n \geq 0\}$ is adapted if:

$$X_n \in \mathcal{F}_n$$

In Adaptation, the data X_n only depends on information until instant n .

Definition 5.3. We consider a sequence of random variables $X = \{X_n; n \geq 0\}$ such that

- $\{X_n; n \geq 0\}$ is adapted.
- $X_n \in L^1(\Omega)$ for all $n \geq 0$.

Then

- X is a martingale if $X_n = \mathbb{E}[X_{n+1} | \mathcal{F}_n]$.
- X is a supermartingale if $X_n \geq \mathbb{E}[X_{n+1} | \mathcal{F}_n]$.
- X is a submartingale if $X_n \leq \mathbb{E}[X_{n+1} | \mathcal{F}_n]$.

We see that the definition of a martingale was shared in definition 1.1, this is the martingale property. This equation emphasizes the current value, X_n is the expected value of the next step, given the information up to n . However,

this definition is in discrete time, meaning that it follows the adaptation and integrability condition. To better understand this idea, we can use an example of the fair game shared in the introduction. Using the same rules, we can now prove this definition.

Example. Let X_n represent the player's total winnings after n games. We will show that $(X_n)_{n \geq 0}$ is a martingale.

Filtration:

Define the filtration $\{\mathcal{F}_n\}_{n \geq 0}$ where \mathcal{F}_n represents the information up to the n th game. Specifically, \mathcal{F}_n contains the outcomes of the first n coin tosses.

Martingale Property:

To prove that $\{X_n\}_{n \geq 0}$ is a martingale with respect to $\{\mathcal{F}_n\}_{n \geq 0}$, we need to verify the three conditions:

- Adaptation: X_n is \mathcal{F}_n -measurable because the total winnings after n games depend only on the outcomes of the first n games.
- Integrability: The expectation $\mathbb{E}[|X_n|]$ is finite. Since X_n is the sum of n independent bets of ± 1 , its absolute value grows linearly with n , ensuring finite expectation.
- Martingale Property:

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

Proof. Given that X_n is the player's total winnings after n games, the total winnings after $n + 1$ games, X_{n+1} , can be written as:

$$X_{n+1} = X_n + Y_{n+1}$$

where Y_{n+1} is the result of the $(n + 1)$ th game

- $Y_{n+1} = 1$ if the player wins the $(n + 1)$ th game (heads).
- $Y_{n+1} = -1$ if the player loses the $(n + 1)$ th game (tails).

The expected value of Y_{n+1} given \mathcal{F}_n is:

$$\mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] = 0$$

because the coin toss is fair, and the expected outcome of a fair coin toss is zero (since

$$P(Y_{n+1} = 1) = P(Y_{n+1} = -1) = 1/2)$$

Therefore,

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = \mathbb{E}[X_n + Y_{n+1} | \mathcal{F}_n] = X_n + \mathbb{E}[Y_{n+1} | \mathcal{F}_n] = X_n + 0 = X_n$$

This confirms that the process $\{X_n\}_{n \geq 0}$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}_{n \geq 0}$

From this example, we can move to another stochastic process like martingales, stopping time.

Definition 5.4. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $\{\mathcal{F}_t\}_{t \geq 0}$, a random variable $\tau : \Omega \rightarrow [0, \infty]$ is a stopping time if for every $t \geq 0$,

$$\{\tau \leq t\} \in \mathcal{F}_t.$$

Lemma 5.5. Let T be a stopping time. If there exists a not random variable, N and $\varepsilon > 0$ such that $\mathbb{P}(T \leq n + N | \mathcal{F}_n) > \varepsilon \quad \forall n \geq 0$, then $\mathbb{E}[T] < \infty$.

Proof. $\mathbb{P}(T > kN) = \mathbb{P}(T > kN \cup T > (k-1)N)$

$$= \mathbb{P}(T > kN | T > (k-1)N) * \mathbb{P}(T > (k-1)N)$$

$$\leq (1 - \varepsilon) * \mathbb{P}(T > (k-1)N) \leq (1 - \varepsilon) * \mathbb{P}(T > (k-2)N) \leq \dots \leq (1 - \varepsilon)^k$$

$$\mathbb{E}[T] = \sum_{l=0}^{\infty} \mathbb{P}(T > l) \leq N \sum_{k=0}^{\infty} \mathbb{P}(T > kN) \leq N \sum_{k=0}^{\infty} (1 - \varepsilon)^k = \frac{N}{1 - (1 - \varepsilon)} = \frac{N}{\varepsilon} < \infty$$

This shows that the expectation of a T , a stopping time is finite. Now, we can move into the martingale transform.

Definition 5.6. We assume we have a probability space with a filtration. C_n is predictable if C_n is \mathcal{F}_{n-1} measurable $\forall n$. Imagine that C_n is the strategy that you are making at time n , but to make prediction at time n , we used the information at time $n-1$. Once we have this base information, we can move to the martingale transform of X .

$$Y_n = \sum_{1 \leq k \leq n} C_k * (X_k - X_{k-1})$$

Theorem 5.7. If C_n is predictable, and $0 \leq C_n \leq k$. $\forall (n, \omega)$ and X is a supermartingale, then so is the martingale transform, $C \cdot X$. If C_n is predictable, and $|C_n| \leq K \forall (n, \omega)$ and X is a martingale, then so is $C \cdot X$

Remark 5.8. If C_n has $\mathbb{E}[C_n^2] < \infty$, and $\mathbb{E}[X_n^2] < \infty \forall n$, then Theorem 5.7 still works without the bound, $|C_n| \leq K$.

Proof. To prove Y is a martingale, then we must prove the following equation: $\mathbb{E}[(Y_n - Y_{n-1})\mathcal{F}_{n-1}] = 0$. To prove this, we must find what the difference of $(Y_n - Y_{n-1})$ is and it should equal $C_n * (X_n - X_{n-1})$, where C_n is \mathcal{F}_{n-1} measurable. Let's begin by factoring out the original equation.

$$\begin{aligned} &= C_n \mathbb{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}) \\ &= C_n (\mathbb{E}(X_n | \mathcal{F}_{n-1}) - \mathbb{E}(X_{n-1} | \mathcal{F}_{n-1})) \end{aligned}$$

If X is a martingale, $X_{n-1} = 0$, which is proved here.

Theorem 5.9. *Let T be a stopping time, if X is a supermartingale, then so is X^T .*

Proof. $C_n := \mathbb{1}(n \leq T) = \mathbb{1}(n - 1 < T)$, which is \mathcal{F}_n measurable.

$$\begin{aligned} (C \cdot X)_n &= \sum_{k=1}^n C_k (X_k - X_{k-1}) \\ &= \begin{cases} T \geq n : \sum_{k=1}^n (x_k - x_{k-1}) = x_n - x_0 \\ T < n : \sum_{k=1}^T (x_k - x_{k-1}) = x_T - x_0 \end{cases} \\ &= X_{T \wedge n} - X_0 = X_n^T - X_0 \end{aligned}$$

Now we can move to the Optional Stopping Theorem, the relation between stopping time and martingales.

Remark 5.10. If M_n is a martingale, then $\mathbb{E}[M_n] = \mathbb{E}(\mathbb{E}[M_n | \mathcal{F}_{n-1}]) = \mathbb{E}[M_{n-1}] = \dots = \mathbb{E}[M_0]$ (from the Tower Property).

Remark 5.11. If T is a stopping time, M_n is a martingale, then $\mathbb{E}[M_{T \wedge n}] = \mathbb{E}[M_0]$.

Theorem 5.12. *Doob's Optional Stopping: Let T be a stopping time, X is a supermartingale. If it follows these conditions:*

- T is bounded OR X is bounded and T is infinite almost surely OR $\mathbb{E}[T] < \infty$ and $|X_n - X_{n-1}| \leq k$

Then $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$

Proof. If T is not almost surely bounded, approximate it by a sequence of bounded stopping times $T_n = \min(T, n)$. Then we can use the martingale property to show that for each T_n , $\mathbb{E}[X_{T_n}] = \mathbb{E}[X_0]$. Finally, apply the Dominated Convergence Theorem to take the limit as $n \rightarrow \infty$ and conclude that $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

After talking a bit about convergence in supermartingales, we can move to understanding Doob's Forward Convergence Theorem.

Definition 5.13. $X \in L_p$, if $\mathbb{E}[|X|^p] < \infty$; $|X|^p := (\mathbb{E}[|X|^p])^{1/p}$, where p is positive.

Definition 5.14. X_n is bounded in L^p , if we take $\sup_n \|X_n\|_p < \infty$. This means if we take the p -norms, it will still be bounded regardless if we take the largest value of n . However, there is a uniform bound, $\|X_n\|_p \leq k < \infty$ for all n .

Lemma 5.15. Using Doob's upcrossing Lemma, if X_n is a supermartingale, then $(b - a)\mathbb{E}[U_n]$ on the interval $[a, b] \leq \mathbb{E}[X_n - a]^-$, we can do a corollary.

Corollary 5.16. Let X be a supermartingale, bounded in L^1 : $\sup_n \mathbb{E}[|X_n|] < \infty$. If we fix $a < b$, then $(b - a)\mathbb{E}[U_\infty]$ on the interval $[a, b] \leq |a| + \sup_m \mathbb{E}[|X_m|] < \infty$. In particular, $\mathbb{P}(U_\infty[a, b] = \infty) = 0$.

Proof. $(b - a)\mathbb{E}[U_n][a, b] \leq \mathbb{E}[X_n - a]^- \leq \mathbb{E}|X_n - a| \leq \mathbb{E}|X_n| + |a| \leq \sup_n \mathbb{E}|X_n| + |a|$. $U_n[a, b]$ is defined as number of upcrossings until time N , so if we know that it is non-decreasing in N , then it has a limit. $\lim_{n \rightarrow \infty} U_n[a, b] = U_\infty[a, b]$. We need to make sure that this expression is finite, proving the corollary. To begin this, we take $(b - a)\mathbb{E}[U_\infty][a, b]$ then use the monotone convergence theorem to find the limit. This means that it equals to $(b - a)\lim_{n \rightarrow \infty} \mathbb{E}U_n[a, b]$, which is $\leq k + |a|$ due to the first statement of the proof.

Now, we can move to what Doob's forward convergence theorem states.

Theorem 5.17. Doob's forward convergence: Let X be a supermartingales, bounded in L^1 . Then $X_\infty := \lim_n X_n$ exists almost surely and is almost surely finite.

Proof. We want to prove that X_n does not convergence. To prove that $X_\infty := \lim_n X_n$ exists almost surely, we begin that $= \bigcup_{a < b} \liminf X_n < a < b < \limsup X_n \subseteq \bigcup_{a < b} U_\infty[a, b] = \infty$, so the probability of this latter part = 0.

$$\mathbb{P}(X_n \text{ does not converge}) = 0$$

For the second part of the theorem, of "almost surely finite," we state that $\mathbb{E}|X_\infty| = \mathbb{E}[\liminf |X_n|] \leq$ (using Fatou's Lemma) $[\liminf] \mathbb{E}|X_n| \leq \sup \mathbb{E}|X_n| < \infty$. Therefore, knowing that X is a supermartingale is bounded in L^1 , we get $|X_\infty| < \infty$ almost surely.

Remark 5.18. If $X_n \geq 0$ is a supermartingale, then the L^1 bound is automatic: $\mathbb{E}|X_n| = \mathbb{E}[X_n] \leq \mathbb{E}[X_0] < \infty$

From here we are able to understand convergence in L^p , but take a case of a L^2 martingale. L^2 martingales: $M_n \in L^2 \quad \forall n$.

Then $M_v - M_u \perp L^2(\mathcal{F}_u)$ In this situation, what does orthogonal exactly mean? Let's assume we have a square integrable random variable that is finite, we can define a similar scalar product. $\mathbb{E}(X \cdot Y) = \langle x, y \rangle$ In the sense that these random variable are almost like vectors, we can fully understand what orthogonality means. Going back to the first first expression, $\forall v > u > t > s$, then following is true, $\mathbb{E}[(M_t - M_s) * (M_v - M_u)] = 0$ This is saying that the increment of my martingale from s to t and u to v are orthongal in the sense that my scalar product is 0. The proof of this is the following.

Proof. $\mathbb{E}[\mathbb{E}[(M_t - M_s) * (M_v - M_u)] | \mathcal{F}_u]$ Because $v > u > t > s$, $M_t - M_s$ are \mathcal{F} -measurable. That means our new expression is, $\mathbb{E}[(M_t - M_s) * \mathbb{E}[(M_v - M_u)] | \mathcal{F}_u]$ And by the martingale property, our latter expectation equals 0, this implies a theorem.

Theorem 5.19. *An L^2 -martingale M_n is bounded in L^2 if. $\sum_{k=1}^{\infty} \mathbb{E}(M_k - M_{k-1})^2 < \infty$. In this case, $M_n \rightarrow M_0$ almost surely in L^2 .*

Proof. We can show proof of the Pythagorean Theorem. $\mathbb{E}[M_n^2] = \|M_n\|_2^2$

$$= \mathbb{E}[M_0 + \sum_{k=1}^{\infty} (M_k - M_{k-1})]^2$$

$$= \mathbb{E}[M_0^2] + \sum_{k=1}^n \mathbb{E}[M_k - M_{k-1}]^2 + 0$$

Therefore, M is bounded in

$$L^2 : \mathbb{E}[M_n^2] = \mathbb{E}[M_0^2] + \sum_{k=1}^n \mathbb{E}[M_k - M_{k-1}]^2 \leq \mathbb{E}[M_0^2] + \sum_{k=1}^{\infty} \mathbb{E}[M_k - M_{k-1}]^2 < \infty$$

Also, bounded in L^1 because of Doob's Forward Convergence, meaning that

$$M_n \xrightarrow{\text{almost surely}} M_{\infty}$$

The second norm states the following.

$$\mathbb{E}[M_{\infty} - M_n]^2 \leq \liminf \mathbb{E}[M_{n+r} - M_n]^2 = \liminf \sum_{k=n+1}^{n+r} \mathbb{E}[M_k - M_{k-1}]^2 = \sum_{k=n+1}^{\infty} \mathbb{E}[M_k - M_{k-1}]^2$$

and as n approaches ∞ , this expression equals 0.

6. APPLICATIONS IN FINANCE

This martingale theory can be applied to stock market strategies. For instance, the Martingale Strategy suggests doubling your investment when you incur a loss. In the context of the stock market, if you purchase a stock at 10 dollars and its price drops to 9 dollars, you would buy another share at the new price. If the stock price falls further to 8 dollars, you would buy two additional shares. Eventually, when the stock price rebounds to 10 dollars, you not only recover your initial investment but also make a profit from the shares purchased at the lower prices. This approach reflects the martingale theory's essence, where the future expected value of the process remains equal to the present value, despite short-term fluctuations.

7. ACKNOWLEDGEMENTS

I want to thank my mentor, Paulina Paiz for the valuable feedback and tips throughout this process. I would like to also thank Simon Rubenstein-Salzedo for organizing the Euler Circle.

REFERENCES

- [1] Ishaan Goel, *INTRODUCTION TO MARTINGALES WITH AN APPLICATION IN FINANCE*. UChicago, 2022.
- [2] Sheffield, *Martingales, risk neutral probability, and Black-Scholes option pricing*. MIT.
- [3] Dr. Samy Tindel, *Martingales*. Purdue University.
- [4] David Williams, *Probability with Martingales*. Cambridge University Press, 1991.

MILTON ACADEMY, MILTON, MA, 02186

Email address: rohan14shah@gmail.com