INTRODUCTION TO MARTINGALES WITH AN APPLICATION IN FINANCE

ISHAAN GOEL

Abstract. Martingales are an important construct in probability theory, which find great utility in mathematically modelling many real-life processes. This paper introduces martingales in a rigorous measure-theoretic manner, along with some interesting properties. Then, we discuss an application of martingales in the realm of finance - specifically the behavior of futures prices. We assume elementary knowledge of measure theory and probability.

Contents

1. Introduction

Consider a gambler playing a game with a fair coin, where he wins \$1 if the coin flips heads, and loses \$1 if the coin flips tails. Assume, after flipping the coin a certain number of times, that the gambler's fortune (his total winnings) is $\$W$. We permit W to be negative (a net loss) for now. After the next flip, there is an equal chance of his fortune reaching $\$W + 1$ and $\$W - 1$. Basic understanding of probability shows that on average, we can expect his fortune to stay $\$W$ - that is, the expected change in his fortune is nil. So, there are no risk or rewards associated with playing this game in the sense that his expected winnings in each round are zero - neither positive nor negative.

This is a simple example of a *fair game*. The gambler does not expect to make a gain or a loss from turn to turn. This is true regardless of how he has been performing in past turns. In the example above, he could have been flipping heads

Date: AUGUST 23, 2022.

for several consecutive turns, but his probability of flipping heads or tails on the *next* turn stays the same. Therefore, using the reasoning above, the expected change in his fortune would remain 0, regardless of past performance. In other words, even given the *information* he gained from past turns, his expected fortune after the next turn would be his current fortune itself.

The probabilistic model of a fair game is called a *martingale*. A martingale process does not necessarily need to be a "game" in the conventional sense. Any random process, where your expectation for the next "value" of the process is the current value, given all information about past values, can be modeled using martingales. To access the mathematical definition of martingales, we assume the reader has elementary knowledge of measure theory (particularly measurable spaces, σ-algebras and Borel sets), as well as an intuitive understanding of probabilistic terms like events, outcomes and expectations.

2. Probability and measure theory

We begin by framing some probability concepts in measure-theoretic terms. Consider an arbitrary random experiment with some outcomes. Let Ω represent the set of all possible outcomes, or the *sample space*, of the experiment. For instance, in the case of rolling a die, we have $\Omega = \{1, 2, 3, 4, 5, 6\}$. An *event* here is defined as any subset of Ω. In the case of the die, the event of "rolling an even number" is the subset $\{2, 4, 6\}$. We now want to assign a *probability* to each possible event, which can be viewed as a *measure* assigned to the subset representing the event.

The next step is to characterize the exact domain of sets to which we want to assign this probability measure. Assume that we have already assigned a measure to two events A and B . Then, we would also want a measure for the events of A and B not occurring, given by the sets A^c and B^c, and for at least one of the events occurring, given by the set $A \cup B$.

This is reminiscent of the definition of a σ-algebra. In particular, if A and B are elements of some σ-algebra $\mathcal{F}$, then the sets A^c, B^c and $A \cup B$ are also elements of the σ-algebra $\mathcal{F}$. Thus, we want our probability measure to be assigned to some σ-algebra created by base events that we are interested in measuring. This motivates the following definition.

Definition 2.1. Let F denote a set of subsets of Ω. The σ-algebra *generated* by F, given by $\sigma(F)$, is the smallest σ-algebra containing F. That is, it has the following properties:

(i) $F \subseteq \sigma(F)$;
(ii) For all σ-algebras Σ on Ω such that $F \subseteq \Sigma$, $\sigma(F) \subseteq \Sigma$.

When Ω is finite, we can define the set of all possible events by its power set 2^Ω and assign a suitable probability measure. However, if Ω is infinite, it might not be possible to assign a measure to every possible subset. The definition above is useful in such cases, as we can create a family of *events of interest* and consider only the σ-algebra generated by this family when assigning a probability measure. This σ-algebra can be thought of as all the information about an experiment contained in some events of interest.

We now formally define the probability measure.

Definition 2.2. Let Ω be a sample space and $\mathcal{F}$ be a σ-algebra on Ω generated by a family of events F. A *probability measure* P is a measure on $(\Omega, \mathcal{F})$ such that $P(\Omega) = 1$. The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*.

From the definition, we see that $P(A)$, $A \in \mathcal{F}$ represents the probability of an event A occurring. The probability of the experiment itself occurring is always 1, represented by the condition $P(\Omega) = 1$. Furthermore, as per the definition of a measure, the probability of a countable union of mutually exclusive events (with no overlapping outcomes) must equal the sum of the individual probabilities of those events. Note that a non-empty event may also have probability 0.

Another important concept is that of a random variable. This is a variable which takes on a numerical value based on the outcome of a random experiment. A simple example is the case of a coin toss, where $\Omega = \{H, T\}$ where H represents heads and T represents tails. Then, we can define a corresponding random variable X as

$$X = \begin{cases} 1 & \text{if heads} \\ 0 & \text{if tails} \end{cases}$$

A random variable may be thought of as a function from the sample space to the real numbers (or in general, any measurable space). It assigns a numerical value to each outcome in the sample space. This motivates the following definition.

Definition 2.3. Consider a probability space $(\Omega, \mathcal{F}, P)$ and a function $f : \Omega \to \mathbb{R}$. The function f is called *measurable* on $\mathcal{F}$ (or $\mathcal{F}$-measurable) if for all Borel sets B in the Borel space $(\mathbb{R}, \mathcal{B})$, we have that $f^{-1}(B) \in \mathcal{F}$. Such a measurable function on the probability space is called a *random variable*.

The pre-image of every open set in $\mathbb{R}$ under the random variable is an event in the σ-algebra for which a probability measure is defined. In particular, this includes all open intervals in $\mathbb{R}$. Many sample spaces have infinitely or uncountably many outcomes, in which the probability of every individual outcome is extremely small or even 0. Therefore, rather than focusing on individual outcomes, it is more helpful to consider the probability that a random variable takes on a value in an interval. This defines a measure on the Borel space $(\mathbb{R}, \mathcal{B})$, related to the measure on the original sample space.

Definition 2.4. Let X be a random variable on $(\Omega, \mathcal{F}, P)$. The *distribution* of X is the function $\mu_X : \mathcal{B} \to [0, 1]$ such that

$$\mu_X(B) = P(X^{-1}B)$$

for all Borel sets $B \in \mathcal{B}$.

Proposition 2.5. *The distribution μ of a random variable X forms a probability measure on* $(\mathbb{R}, \mathcal{B})$.

Measurable functions satisfy a number of important properties that will be of use to us later. For the sake of avoiding lengthy discussions of measure theory, proofs will not be provided, but may be found in [15, ch. 3].

Theorem 2.6. *Let $f, g : \Omega \to \mathbb{R}$ be measurable functions on the probability space $(\Omega, \mathcal{F}, P)$. Then, f and g satisfy the following properties:*

(i) $f + g$ and $f \cdot g$ are measurable;
(ii) For any $\lambda \in \mathbb{R}$, λf and λg are measurable;

(iii) Any finite combination of the above two is measurable;
(iv) The pointwise maximum of f and g, $\max\{f(\omega), g(\omega)\}\ \forall\, \omega \in \Omega$, *is measurable.*

We now consider the equality of two random variables. It is possible that two random variables might not map to the same values for certain outcomes. However, if the set of outcomes where they differ has probability 0, then they are equal with probability 1. In simpler terms, the differences between the functions are negligible. This notion of equality is formalized as follows.

Definition 2.7. Let $f : \Omega \to \mathbb{R}$ and $g : \Omega \to \mathbb{R}$ be measurable functions on the probability space $(\Omega, \mathcal{F}, P)$. Let $A = \{\omega \in \Omega \mid f(\omega) \neq g(\omega)\}$. If $P(A) = 0$, then we say the functions are equal *almost surely (a.s.)* or *with probability 1 (w.p.1)*.

3. Lebesgue Integration

Given a random variable X, we know it can take on a wide range of values, but we would like to understand its "average" value. That is, the single value that best approximates the distribution of X in $\mathbb{R}$. This is called the *expectation* or *expected value* of the random variable.

If X takes on a finite number of values, the expectation is found simply by taking the average of all possible values weighted by their probability. For instance, suppose that $X = 1$ with probability $\frac{1}{2}$ and $X = -1$ with probability $\frac{1}{2}$. Then the expectation of X, represented as $\mathbb{E}(X)$, is $1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$.

If X takes on an infinite number of values, then finding the weighted average would require integration to evaluate this infinite sum. However, the usual Riemann integral is often unsuitable in this context. The Riemann integral involves splitting the domain into contiguous intervals to form "rectangles", and approximates the integral through the convergent sum of areas of these rectangles. However, this notion of intervals and rectangles might not make sense for an abstract sample space. Furthermore, some measurable functions may not satisfy the continuity requirements necessary for Riemann integration. Therefore, we utilize a generalized form of the Riemann integral called the *Lebesgue integral.* To understand Lebesgue integration, we first introduce simple functions in a general, non-probabilistic sense.

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

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

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 2.2, $\mathbf{1}_A$ is measurable. $\square$

Certain measurable functions can be completely expressed as linear combinations of indicator functions. Consider a measurable function f where the sample space can be divided into finitely many disjoint subsets, where all outcomes in a single subset return the same value under f. Thus, f takes on finitely many values, and we get the following definition.

Definition 3.3. Consider the measure space $(\Omega, \mathcal{F}, \mu)$. A measurable function $f : \Omega \to [0, \infty]$ is called *simple* if there exist disjoint subsets $\{A_1, A_2, A_3, \ldots, A_k\}$ of Ω and corresponding reals $\{a_1, a_2, \ldots, a_k\}$ such that

$$f = \sum_{i=1}^{k} a_i \mathbf{1}_{A_i}.$$

The Lebesgue integral of f is defined as

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

We will write $\int_{\Omega}$ as $\int$ when there is no confusion about the domain over which we are integrating.

The utility of the Lebesgue integral becomes more apparent with this definition. The Lebesgue integral of a simple function is simply the average of its possible values weighted by the measure of the subset of the domain giving this value. 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 giving those values - that is, the expectation.

For non-simple functions, the integral may be defined by approximating the function through simple functions. This requires the following theorem.

Theorem 3.4. *Let $f : \Omega \to [0, \infty]$ 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. Its proof is beyond the scope of this paper, but may be found at [4]. We can then define the Lebesgue integral of an arbitrary non-negative function as follows.

Definition 3.5. Let $f : \Omega \to [0, \infty]$ be a measurable function on the space $(\Omega, \mathcal{F}, \mu)$, and let S be the set of all simple functions s such that $s(\omega) \leq f(\omega) \; \forall \; \omega \in \Omega$. Then, the Lebesgue integral of f is defined as

$$\int f d\mu = \sup \left\{ \int s d\mu \right\}_{s \in S}.$$

The Lebesgue integrals of non-negative measurable functions satisfy several nice properties, but we only need one for our purposes.

Proposition 3.6. *For any two non-negative measurable functions $f, g : \Omega \to \mathbb{R}$,*

$$\int (f + g) d\mu = \int f d\mu + \int g d\mu.$$

The proof for this proposition is deceptively complex, and may be found at [2]. Functions that are not non-negative can be represented by a sum of functions that are, as shown in the following lemma.

Lemma 3.7. *Let $f : \Omega \to \mathbb{R}$ be any measurable function on $(\Omega, \mathcal{F}, \mu)$. We define*

$$f^+(x) = \max\{f(x), 0\},$$
$$f^-(x) = \max\{-f(x), 0\}.$$

Then $f^+(x)$ and $f^-(x)$ are non-negative and measurable, and $f(x) = f^+(x) - f^-(x)$.

Proof. The non-negativity of the functions is trivial to check. We note that any constant function is measurable. This is because the pre-image of a Borel set containing the constant is Ω, and the pre-image of a Borel set not containing the constant is $\emptyset$, both of which are contained in any σ-algebra. By Theorem 2.6 (ii), the negation of any measurable function is measurable. Then, by Theorem 2.6 (iv), we see that $f^+(x)$ and $f^-(x)$ are pointwise maxima of two measurable functions and are therefore themselves measurable. It is trivial to check that $f(x) = f^+(x) - f^-(x)$. $\square$

Having proved this lemma, we can use Definition 3.5 to find the Lebesgue integrals of $f^+(x)$ and $f^-(x)$, and combine them using Proposition 3.6 to produce the following definition.

Definition 3.8. Let $f : \Omega \to \mathbb{R}$ be any measurable function on the space $(\Omega, \mathcal{F}, \mu)$. Then, its Lebesgue integral is defined as

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

This definition only holds when at least one of $\int f^+ d\mu$ and $\int f^- d\mu$ is finite. Not all measurable functions have a finite Lebesgue integral, so it is convenient to specially designate the ones that do.

Definition 3.9. A measurable function $f : \Omega \to \mathbb{R}$ on $(\Omega, \mathcal{F}, \mu)$ is said to be *Lebesgue integrable* if

$$-\infty < \int_\Omega f d\mu < \infty.$$

In other words, its integral over the entire domain is finite.

It might be difficult to verify the integrability of some functions from the definition above, so we prove below an easier equivalent condition for integrability.

Definition 3.10. The *absolute value* of a measurable function $f : \Omega \to \mathbb{R}$ is defined as

$$|f| = \begin{cases} f(\omega) & \text{for} \quad f(\omega) \geq 0 \\ -f(\omega) & \text{for} \quad f(\omega) < 0 \end{cases}.$$

Theorem 3.11. *A measurable function $f : \Omega \to \mathbb{R}$ is integrable if, and only if, its absolute value is integrable.*

Proof. Note that $|f| = f^+(x) + f^-(x)$ - this is trivial to check. Then, using the same reasoning as for f, we have that $\int |f| d\mu = \int f^+ d\mu + \int f^- d\mu$. With this, we now prove both directions.

($\Longrightarrow$) Assume f is integrable. Then, $-\infty < \int f < \infty \implies -\infty < \int f^+ - \int f^- < \infty$. Since their difference is finite, $\int f^+$ and $\int f^-$ must be finite and therefore $\int f^+ + \int f^- = \int |f|$ is also finite. Thus, $|f|$ is integrable.

($\Longleftarrow$) Assume $|f|$ is integrable. Then, $-\infty < \int f^+ + \int f^- < \infty$ and so $\int f^+$ and $\int f^-$ are finite. Clearly, their difference must also be finite and therefore $-\infty < \int f^+ - \int f^- = \int f < \infty$. Thus, f is integrable. $\square$

A final thing to discuss is the Lebesgue integral over restricted domains. This is conceptually similar to restricting the domain for the Riemann integral. Intuitively, if we were restricting our domain of integration to some subset A of Ω, we would simply want the function to "vanish" at any points not in the target domain. This is achieved as follows.

Definition 3.12. For any measurable $A \subseteq \Omega$ and measurable function $f : \Omega \to \mathbb{R}$ on the space $(\Omega, \mathcal{F}, \mu)$, we define the integral of f on A as

$$\int_A f d\mu = \int_\Omega \mathbf{1}_A f d\mu.$$

We conclude this section with a theorem related to integration on restricted domains. Its proof is excluded due to complexity, but may be found by referring to [1] and [11].

Theorem 3.13. *Let $f, g : \Omega \to \mathbb{R}$ be measurable functions on the probability space $(\Omega, \mathcal{F}, \mu)$. Then $f = g$ almost everywhere if, and only if, $\int_A f d\mu = \int_A g d\mu$ for all $A \in \mathcal{F}$.*

4. Conditional Expectation

Now that we have introduced the Lebesgue integral, the expectation of a random variable can be defined directly as follows.

Definition 4.1. The *expectation* of a random variable X on the probability space $(\Omega, \mathcal{F}, P)$ is defined as $\mathbb{E}(X) = \int_\Omega X dP$.

Sometimes, we might have information that might help us approximate a random variable better, by restricting the values that it could take. Our expected value of the random variable would then change based on the information we have about it.

For example, consider a simple experiment where two dice are rolled in succession. Let the values on the faces of the dice be a and b. Then, let X be the random variable representing the sum of the numbers on the die faces - $X = a + b$. X can take any integer value between 2 and 12, with varying probabilities. We can calculate, using the weighted sum approach discussed at the beginning of the previous section, that $\mathbb{E}(X) = 7$.

Now suppose that we know that $a = 5$ - that is, the first roll returned a 5. The possible values of X are now restricted to integers from 6 to 11, with equal probabilities of $\frac{1}{6}$. Thus, the expected value of X with this new information is $6 \cdot \frac{1}{6} + 7 \cdot \frac{1}{6} + 8 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 10 \cdot \frac{1}{6} + 11 \cdot \frac{1}{6} = 8.5$. Notice that both the possible values of X and their associated probabilities changed.

We see that the information received serves as a subset of the overall sample space - in other words, an event. If we know an event has occurred, we know that its complementary event has not occurred and has probability 0. In integral terms, the random variable function effectively "vanishes" over outcomes not in this event, which changes the associated probability of each outcome. This is equivalent to integrating the random variable over a restricted domain.

We could also gain information from more than one event. Say we know that two events A and B in the sample space have occurred. We gain the following information from this knowledge:

(i) $\emptyset$ has not occurred and Ω has occurred
(ii) the complementary events A^c and B^c have not occurred
(iii) the event $A \cup B$ has occurred (this is true even if one of A or B occur)

Let $\mathcal{G}$ be the σ-algebra generated by A and B. Then, we notice that the sets $\emptyset, \Omega, A^c, B^c, A \cup B \in \mathcal{G}$. Therefore, $\mathcal{G}$ effectively represents all of the information we have about the outcomes of the experiment. In general, if we know a number of events have occurred, then the information obtained from those events is captured in a σ-algebra $\mathcal{G}$ generated by that family of events. The probability of some events within that σ-algebra may be 0, but this still counts as information about their occurrence.

$\mathcal{G}$ would always be a subset of the σ-algebra $\mathcal{F}$ of the probability space, generated by all possible events. Our expectation of X changes based on the events contained in $\mathcal{G}$, or is *conditioned* on $\mathcal{G}$. We then have the following definition.

Definition 4.2. Let X be an integrable random variable on the probability space $(\Omega, \mathcal{F}, P)$, and $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra on Ω. The *conditional expectation* of X given $\mathcal{G}$ is an integrable, $\mathcal{G}$-measurable random variable $\mathbb{E}(X \mid \mathcal{G})$, such that for all events E in $\mathcal{G}$, we have

$$\int_E \mathbb{E}(X \mid \mathcal{G}) dP = \int_E X dP.$$

Definition 4.3. A random variable Y such that $Y = \mathbb{E}(X \mid \mathcal{G})$ almost surely is called a *version* of $\mathbb{E}(X \mid \mathcal{G})$.

Proposition 4.4. *All random variables satisfying Definition 4.2 are versions of* $\mathbb{E}(X \mid \mathcal{G})$. *That is, the conditional expectation is unique up to almost sure equivalence.*

The proof for this proposition follows directly from Theorem 3.13. Expectations and conditional expectations both satisfy the property of linearity, which is presented in the theorem below without proof.

Theorem 4.5. *Let* X, Y *be random variables on the probability space* $(\Omega, \mathcal{F}, P)$, *and let* $\mathcal{G} \subseteq \mathcal{F}$ *be a* σ*-algebra. Then, the following statements are true:*

(i) $\mathbb{E}(X) + \mathbb{E}(Y) = \mathbb{E}(X + Y)$;
(ii) If $\mathbb{E}(X)$, $\mathbb{E}(Y)$ *and* $\mathbb{E}(X+Y)$ *all exist, then* $\mathbb{E}(X \mid \mathcal{G}) + \mathbb{E}(Y \mid \mathcal{G}) = \mathbb{E}(X + Y \mid \mathcal{G})$;
(iii) For all $\alpha \in \mathbb{R}$, $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$ *and* $\mathbb{E}(\alpha X \mid \mathcal{G}) = \alpha \mathbb{E}(X \mid \mathcal{G})$.

5. Martingales

We now return to the concept of martingales and fair games. We can consider the "value" of a fair game, such as the fortune of the gambler from the coin-flipping game originally discussed, as a time-indexed sequence of random variables. That is, the value of the game at some time t would be given by a random variable X_t, that would change in the next time-step $t + 1$.

In the first section, we explained that the expectation for the value of the game in future times would be the current value of the game, regardless of the information we have about the game up till that point. We can think of the information we possess about a game up to and including time t as a σ-algebra $\mathcal{F}_t$. The information we have grows at each time-step as more events occur and outcomes are realized. We now present a formal definition of martingales using this terminology.

Definition 5.1. Let $\{X_n\}_{n\in\{1,2,3,\dots\}}$ be a sequence of random variables on the probability space $(\Omega, \mathcal{F}, P)$. Let $\{\mathcal{F}_n\}_{n\in\{1,2,3,\dots\}}$ be a sequence of σ-algebras such that $\mathcal{F}_n \subseteq \mathcal{F}$ for all n. Then, we call the sequence $\{(X_n, \mathcal{F}_n)\}_{n\in\{1,2,3,\dots\}}$ a *martingale* if it satisfies the following conditions:

(i) $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$;
(ii) X_n is measurable with respect to $\mathcal{F}_n$;
(iii) $\mathbb{E}(|X_n|) < \infty$;
(iv) $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$ almost surely.

The first condition establishes that the information we have at each step includes all previous information, along with some new events. This concept is practically called *learning without forgetting*. The sequence $\{\mathcal{F}_n\}_{n\in\{1,2,3,\dots\}}$, when it satisfies property (i), is called a *filtration* on the probability space.

Often, the information we have after a certain time n will just be the σ-field generated by the random variables $X_1, X_2, \dots, X_n$, or by another sequence of random variables $Y_1, Y_2, \dots, Y_n$. This σ-field represents the smallest set of events that we gain information about given past realizations of the stochastic process. We represent this σ-field by $\sigma(X_1, X_2, \dots, X_n)$ or $\sigma(Y_1, Y_2, \dots, Y_n)$ as the case may be. Instead of $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n)$, we would write $\mathbb{E}(X_{n+1} \mid X_1, X_2, \dots, X_n)$ or $\mathbb{E}(X_{n+1} \mid Y_1, Y_2, \dots, Y_n)$.

The second condition is technical and ensures that X_n at every stage is measurable with respect to the σ-algebra representing the information. This condition is often phrased as "X_n is *adapted* to $\mathcal{F}_n$". Note that in the case of the martingale being adapted to previous realizations of a stochastic process, this measurability condition is satisfied by definition of σ-fields.

The third condition, using Definition 4.1, necessitates that the random variable represented by the absolute value of X_n be integrable. Using Theorem 3.11, we see this is equivalent to the condition that X_n be integrable for all n.

The fourth condition is commonly referred to as the *martingale property*. It says that our expectation for X_{n+1}, given all the information we have till time n, is X_n itself. This is the characteristic feature of fair games, where we cannot use past information about the game to guarantee any profits or losses in the next step. This property can be generalized to any number of future steps.

Proposition 5.2. *Let $\{(X_n, \mathcal{F}_n)\}_{n\in\{1,2,3,\dots\}}$ be a martingale. Then, for all $k \in \mathbb{N}$, we have that $\mathbb{E}(X_{n+k} \mid \mathcal{F}_n) = X_n$ almost surely.*

Proof. From Definition 5.1, we have $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$ almost surely. From Theorem 3.13, for any event $E \in \mathcal{F}_n$, we have

$$\int_E \mathbb{E}(X_{n+1} \mid \mathcal{F}_n)\, dP = \int_E X_n\, dP.$$

However, from Definition 4.2, we have

$$\int_E \mathbb{E}(X_{n+1} \mid \mathcal{F}_n)\, dP = \int_E X_{n+1}\, dP.$$

Therefore, we have

$$\int_E X_{n+1}\, dP = \int_E X_n\, dP.$$

If $E \in \mathcal{F}_n$, then $E \in \mathcal{F}_{n+1}$ by Definition 5.1. Then, we can apply the reasoning above on $\mathbb{E}(X_{n+2} \mid \mathcal{F}_{n+1})$ to show that $\int_E X_{n+2} = \int_E X_{n+1} = \int_E X_n$ for all $E \in \mathcal{F}_n$. Applying this process inductively, we get $\int_E X_{n+k} = \int_E X_{n+k-1} = \cdots = \int_E X_{n+1} = \int_E X_n$, since $E \in \mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \cdots \subseteq \mathcal{F}_{n+k-1}$.

Thus, using Definition 4.2, $\int_E X_{n+k} = \int_E \mathbb{E}(X_{n+k} \mid \mathcal{F}_n) = \int_E X_n$ for all $E \in \mathcal{F}_n$, and so by Theorem 3.13 $\mathbb{E}(X_{n+k} \mid \mathcal{F}_n) = X_n$ almost surely. □

This property shows that no matter how far into the future we look, in a fair game our expectation for the value of the game will remain its current value. We can also modify the martingale property slightly to model what an unfair game would look like.

Definition 5.3. A sequence $\{(X_n, \mathcal{F}_n)\}_{n\in\{1,2,3,\dots\}}$ is called a *submartingale* if it satisfies properties (i) - (iii) in Definition 5.1, but

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \geq X_n.$$

It is called a *supermartingale* if

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) \leq X_n.$$

To illustrate, consider our initial coin-flipping game with the gambler and his fortune. His fortune at a turn n can be modelled using a random variable W_n. We showed earlier that the sequence of fortunes is adapted to the previous fortune variables and satisfies the martingale property. We can trivially check that it satisfies the other properties in Definition 5.1, and is therefore a martingale sequence.

However, now assume that the coin is unfair and has a higher probability of returning heads. We can check that the expected fortune at turn $n + 1$ would be higher than the gambler's fortune at turn n, and so it is a submartingale - a game in the player's favor. If the coin had a higher possibility of returning tails, then the opposite would be true and the fortune sequence would be a supermartingale - a game that works against the player.

We now prove another important property of conditional expectations that is relevant to martingales.

Theorem 5.4 (Tower Property)**.** *Let X be an integrable, $\mathcal{F}$-measurable random variable on the probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_m$, $\mathcal{F}_n$ be σ-algebras such that $\mathcal{F}_m \subseteq \mathcal{F}_n \subseteq \mathcal{F}$. Then, we have*

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_n) \mid \mathcal{F}_m) = \mathbb{E}(X \mid \mathcal{F}_m)$$

almost surely.

Proof. By Definition 4.2, we have that for all events $E \in \mathcal{F}_n$,

$$\int_E \mathbb{E}(X \mid \mathcal{F}_n)\, dP = \int_E X\, dP.$$

We also have from Definition 4.2 that for all events $E \in \mathcal{F}_m$,

$$\int_E X \, dP = \int_E \mathbb{E}(X \mid \mathcal{F}_m) \, dP$$

However, by assumption, we have that $\mathcal{F}_m \subseteq \mathcal{F}_n$, therefore $E \in \mathcal{F}_m \implies E \in \mathcal{F}_n$. Thus, we can combine the two equalities above to see that for all $E \in \mathcal{F}_m$,

$$\int_E \mathbb{E}(X \mid \mathcal{F}_n) \, dP = \int_E \mathbb{E}(X \mid \mathcal{F}_m) \, dP.$$

From Definition 4.2, we have that for all $E \in \mathcal{F}_m$,

$$\int_E \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_n) \mid \mathcal{F}_m) \, dP = \int_E \mathbb{E}(X \mid \mathcal{F}_n) \, dP.$$

Combining the equalities above, we get

$$\int_E \mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_n) \mid \mathcal{F}_m) \, dP = \int_E \mathbb{E}(X \mid \mathcal{F}_m) \, dP$$

and so by Theorem 3.13, we get the desired result $\mathbb{E}(\mathbb{E}(X \mid \mathcal{F}_n) \mid \mathcal{F}_m) = \mathbb{E}(X \mid \mathcal{F}_m)$ almost surely. □

6. Application in futures pricing

We now digress briefly from probability theory to introduce an area in finance where we will apply martingales.

6.1. **Introduction to problem and futures pricing.** Financial markets comprise the trade of financial instruments, such as stocks, which are priced competitively based on the laws of supply and demand. A fundamental question that one can ask about the financial markets is whether the prices of these instruments can be predicted using data from the past and present. For example, is it possible to predict the price of a stock based on its current and previous prices, as well as corresponding information about the market?

This is a question of not just academic but practical importance. If there indeed is a method of predicting prices, then this method would be of great commercial value to investors, banks and many other stakeholders. However, if prices are unpredictable given past information, then we are led to the conclusion that no level of expertise can help one profit from the markets any more than an investor holding a random collection of stocks.

In this section, we will follow the works of Paul Samuelson, the "father of modern economics", in [10] to use martingale theory to answer this question for one specific type of financial instrument - futures contracts.

A *futures contract* is an agreement between two parties to buy or sell an asset at a certain time in the future for a certain price [5]. The asset in question might be a commodity such as wheat or gold, stocks of companies, currencies and Treasury bonds, amongst others. Parties generally enter futures contracts when they believe that setting a predetermined *forward price* or *futures price* for the asset would be more profitable than buying or selling the product at its current *spot price* in the market. The time at which the asset purchase occurs is called the *time of maturation*.

For instance, consider a firm that needs to purchase a certain mineral three months in the future. It might expect the spot price of the mineral to rise, or would find the cost of buying the mineral now and holding it for three months too high. It may then enter a futures contract with a futures price lower than the expected spot price of the mineral at the time of maturation, as well as the current price with the added cost of storage. Parties that expect the price of an asset to rise are said to take a *long position* in its market, while those that expect its price to fall take a *short position.*

6.2. **Futures pricing as a stochastic process.** We can model the spot price of an asset as a random variable on some probability space. This random variable would model the possible values that the spot price could take, with an associated probability. Note that the actual distribution of this variable is unknown to us, but using an arbitrary random variable helps us model the asset prices.

Let X_t be the spot price of some asset at time t, and let $\{X_t, X_{t-1}, \dots\}$ be the sequence of all its spot prices up till time t, defined on a probability space $(\Omega, \mathcal{F}, P)$. We consider discrete time-periods like days, months etc. We are interested in futures contracts for some future time T periods after the present time. The spot price then would be X_{t+T}.

We can model futures prices in a similar fashion using random variables. Let $Y(t,T)$ represent the future price of an asset quoted at time t for T periods in the future - that is, a future price for the time of maturation $t+T$. After one time period, the spot price of the asset is X_{t+1}, and the new future price quote for the same time of maturation $t+T$ is now $Y(t+1, T-1)$. Thus, futures prices for a particular time of maturation keep changing every period and are modelled by the sequence $\{Y(t,T), Y(t+1,T-1), \dots, Y(t+T-1, 1), Y(t+T, 0)\}$.

Along each step of the way, we also gain increasing information about the spot prices and market. In particular, we have information about the past realizations $x_0, x_1, \dots$ of the spot price variables $X_0, X_1, \dots$. The information at time t, $\mathcal{F}_t$, is at least as much as we can know from the past spot price variables - in probabilistic terms, we can say that $\mathcal{F}_t \supseteq \sigma(X_t, X_{t-1}, \dots)$. However, we will assume here that $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \dots)$ to simplify our analysis.

6.3. **Martingale property of rationally expected futures prices.** Our aim now is to find a relation between the two sequences $\{Y(t+T-n, n)\}$ and $\{X_n\}$. At the time of maturation, the futures price $Y(t+T, 0)$ is in effect equivalent to the spot price X_{t+T} and cannot differ from it. If it did, then people would buy more of the lower-priced option, raising its price, and sell more of the higher-priced option, lowering its price, until the two balance. Therefore, $Y(t+T, 0) = X_{t+T}$.

For all other time intervals, Samuelson assumes that it is possible to use past data to estimate what the spot price would be T periods from the present. The exact method used to achieve this is not relevant, but he assumes that one could possibly generate a probability distribution for spot prices T periods from the present. Traders would then purchase futures contracts with prices set at the *mean* of this distribution - that is, their best guess for the expected value of the spot price. This leads to the fundamental axiom of his work, given here as a conjecture.

Conjecture 6.1 (Mathematically Expected Price Formation)**.** *Assume that future spot prices may be assigned a probability distribution as described above. Then, for every future time $T = 1, 2, \dots,$ we have that*

$$Y(t,T) = \mathbb{E}(X_{t+T} \mid \mathcal{F}_t).$$

In other words, the futures price is set at the best expectation of the market for the future spot price of the asset. Under this axiom, we can posit the following theorem.

Theorem 6.2. *The sequence of futures prices $\{Y(t+n, T-n)\}$, satisfying Conjecture 6.1 and adapted to the information $\{\mathcal{F}_{t+n}\}$, is a martingale.*

Proof. We will verify all the conditions in Definition 5.1. Condition (i) is satisfied by the definition of $\mathcal{F}_{t+n}$ as a σ-field generated by a growing sequence of variables. Thus, it is always true that $\mathcal{F}_{t+n} \subseteq \mathcal{F}_{t+n+1}$.

From Definition 4.2, we see that $\mathbb{E}(X_{t+T} \mid \mathcal{F}_t)$ is $\mathcal{F}_t$-measurable for all t. By Conjecture 6.1, $Y(t,T)$ is also $\mathcal{F}_t$-measurable. This is true for all $t, t+1, \dots, t+T$, and so condition (ii) is satisfied.

Definition 4.2 also gives us that $\mathbb{E}(X_{t+T} \mid \mathcal{F}_t)$ is integrable and

$$\int_E \mathbb{E}(X_{t+T} \mid \mathcal{F}_t)\, dP = \int_E X_{t+T}\, dP$$

for all $E \in \mathcal{F}_t$. By Conjecture 6.1, we have

$$\int_E Y(t,T)\, dP = \int_E X_{t+T}\, dP.$$

Practically, we know that spot prices are always finite and bounded in some range, so we will assume for ease of analysis that all spot price random variables are integrable. By Theorem 3.11, then, we have that

$$\int_E Y(t,T)\, dP = \int_E X_{t+T}\, dP < \infty.$$

Thus, $Y(t,T)$ is integrable and this is true for all $t, t+1, \dots, t+T$, so condition (iii) is satisfied.

We now prove the martingale property. Specifically, we are to prove that

$$\mathbb{E}(Y(t+1, T-1) \mid \mathcal{F}_t) = Y(t,T).$$

By Conjecture 6.1, we have

$$\mathbb{E}(Y(t+1, T-1) \mid \mathcal{F}_t) = \mathbb{E}(\mathbb{E}(X_{t+T} \mid \mathcal{F}_{t+1}) \mid \mathcal{F}_t).$$

We can check that the conditions of Theorem 5.4 are met here, and apply that to get

$$\mathbb{E}(\mathbb{E}(X_{t+T} \mid \mathcal{F}_{t+1}) \mid \mathcal{F}_t) = \mathbb{E}(X_{t+T} \mid \mathcal{F}_t).$$

Applying Conjecture 6.1 again and combining the equalities, we get the desired result

$$\mathbb{E}(Y(t+1, T-1) \mid \mathcal{F}_t) = Y(t,T).$$

almost surely.

Thus, the sequence satisfies all the conditions in Definition 5.1, and is a martingale. □

The consequence of this finding is that there is no way of making an expected profit using past changes in the spot price. If market players behave in the way that Samuelson hypothesizes they will, then all information they have about the market is already incorporated into the price they set for futures contracts in the current period. There is no additional information that can be exploited to obtain a changed prediction for the futures price in later periods - their best estimate for later futures prices is the current price. This is equivalent to a fair game, as defined in the Introduction section.

6.4. **Risk-free rate and generalizing results.** Samuelson adds to this first theorem by factoring in the *risk-free rate of interest.* There are certain parties in the economy, such as the central bank, that guarantee a risk-free rate of interest. Money invested or borrowed through these parties is guaranteed to be paid or charged a fixed rate of interest. Samuelson proposes that investors would want their futures contract to promise a rate of return equal to the risk-free rate. That is, they would want the future price they pay at the time of maturation to be equivalent to the price they would pay if they earned the risk-free rate on their money the entire time and used that to buy the asset at its future spot price. Differences in the two would cause investors to prefer one option over the other, which would lead to changes in their prices until they equalized.

This reasoning is captured in Samuelson's second axiom, again presented as a conjecture below.

Conjecture 6.3 (Present-Discounted Expected Value)**.** *Let $\lambda = (1 + r)$, where r represents the risk-free rate. Then, for every future time $T = 1, 2, \dots,$ we have that*

$$Y(t, T) = \lambda^{-T}\mathbb{E}(X_{t+T} \mid \mathcal{F}_t).$$

Theorem 6.4. *The sequence of futures prices $\{Y(t + n, T - n)\}$, satisfying Conjecture 6.3 and adapted to the information $\{\mathcal{F}_{t+n}\}$, is a submartingale.*

Proof. We show that the sequence satisfies the four conditions laid out in Definition 5.3. The first three conditions are common to martingales, and can be proved the same way as in Theorem 6.2 with minor changes. We need only show that the submartingale property is satisfied to complete the proof. Specifically, we are to prove that

$$\mathbb{E}(Y(t + 1, T - 1) \mid \mathcal{F}_t) \geq Y(t, T).$$

By Conjecture 6.3, we have

$$\mathbb{E}(Y(t + 1, T - 1) \mid \mathcal{F}_t) = \mathbb{E}(\lambda^{1-T}\mathbb{E}(X_{t+T} \mid \mathcal{F}_{t+1}) \mid \mathcal{F}_t).$$

Applying Theorem 4.5 and Theorem 5.4, we get

$$\mathbb{E}(\lambda^{1-T}\mathbb{E}(X_{t+T} \mid \mathcal{F}_{t+1}) \mid \mathcal{F}_t) = \lambda^{1-T}\mathbb{E}(X_{t+T} \mid \mathcal{F}_t).$$

Applying Conjecture 6.3 again and combining the equalities, we get

$$\mathbb{E}(Y(t + 1, T - 1) \mid \mathcal{F}_t) = \lambda^{1-T}\lambda^{T}Y(t, T) = \lambda Y(t, T).$$

However, $\lambda > 1$, and so $\mathbb{E}(Y(t + 1, T - 1) \mid \mathcal{F}_t) \geq Y(t, T)$. Thus, the sequence satisfies the submartingale property. □

This generalized theorem has a few consequences. It says that we can expect the futures price in the next period to increase by a percentage equal to the risk-free rate, and this is our best guess for the futures price set in all subsequent periods. We cannot use past information to guarantee a profit any greater than the risk-free rate. In stronger terms, this suggests that the markets are not a better investment than risk-free instruments like central bank bonds. It also suggests that the futures price would always be lower than the expected spot price at the time of maturation - a phenomenon that would always hurt the seller of the asset. This phenomenon is called "normal backwardation", and there are various other hypotheses to explain its existence.

Acknowledgements

I would like to thank my mentor, Charley Devlin, for his invaluable support and comments on my paper. I would also like to thank Peter May and the other faculty of the REU for organizing such an enriching experience. Finally, I would like to thank my parents for always enpowering me to pursue the opportunities I want.

References

[1] Eli Bussel. *Martingale approximations for ergodic systems*. `http://math.uchicago.edu/~may/REU2018/REUPapers/Bussel.pdf`. 2018.

[2] The Analysis of Data. *F.3. The Lebesgue Integral*. URL: `http://theanalysisofdata.com/probability/F_3.html#:~:text=Definition%5C%20F.3.5.,%5C%CE%5C%BC%5C%2Dintegrable%5C%20or%5C%20simply%5C%20integrable.`.

[3] Mark Dean. *Measure Theory*. `http://www.columbia.edu/~md3405/DT_Risk_2_15.pdf`. 2015.

[4] Robert Gardner. *3.2 Sequential Pointwise Limits and Simple Approximations - Proofs of Theorems*. `https://faculty.etsu.edu/gardnerr/5210/Beamer-Proofs/Proofs-3-2-print.pdf`. 2018.

[5] John C. Hull. *Options, Futures, and Other Derivatives*. Pearson Education, Inc., 2017.

[6] Gregory F. Lawler. *Introduction to Stochastic Processes*. Chapman & Hall/CRC, 2006.

[7] Burton G. Malkiel. *The Efficient Market Hypothesis and Its Critics*. Tech. rep. 91. Princeton, NJ: Griswold Center for Economic Policy Studies, Apr. 2003.

[8] Anastasios G. Malliaris. "Martingale methods in financial decision-making". In: *SIAM Review* 23.4 (1981), pp. 67–76.

[9] Alessandro Rinaldo. *Conditional Probability and Expectation*. 2018. URL: `https://www.stat.cmu.edu/~arinaldo/Teaching/36752/S18/Notes/lec_notes_6.pdf`.

[10] Paul A. Samuelson. "Proof that properly anticipated prices fluctuate randomly". In: *Industrial Management Review* 6 (1965), pp. 41–49.

[11] Proof Wiki. *A.E. Equal Positive Measurable Functions have Equal Integrals*. `https://proofwiki.org/wiki/A.E._Equal_Positive_Measurable_Functions_have_Equal_Integrals`.

[12] The Free Encyclopedia Wikipedia. *Borel Sets*. `https://en.wikipedia.org/wiki/Borel_set`.

[13] The Free Encyclopedia Wikipedia. *Borel Sets*. `https://en.wikipedia.org/wiki/Random_variable`.

[14] The Free Encyclopedia Wikipedia. *Event (Probability Theory)*. `https://en.wikipedia.org/wiki/Event_(probability_theory)`.

[15] David Williams. *Probability with martingales*. Cambridge University Press, 1991.