AN INTRODUCTION TO ITÔ CALCULUS

FEDERICO GUGLIELMOTTI

Abstract. In this paper, we will introduce the fundamental ideas and prove the main results of Stochastic calculus. We will assume little knowledge of Probability theory and calculus. At the end, we will also look at an important application of stochastic calculus in finance.

Contents

1. Introduction

Stochastic calculus is the area of mathematics that models processes with a stochastic component. The stochastic component is often expressed in terms of Brownian motion, so we will begin this paper by defining it and listing its main properties. We will then define the Itô integral, which allows us to integrate stochastic processes. We will then state and prove Itô's lemma, which is akin to the fundamental theorem of calculus in Riemann calculus. Finally, we will define Stochastic Differential equations and show the existence of solutions under certain assumptions, and we will look at an application of stochastic differential equations in finance; the Black–Scholes equation.

2. Mathematical Preliminaries

2.1. Definitions.

Definition 2.1. A sample space Ω is the set of all possible outcomes of a statistical experiment. For example, the sample space of a coin flip is $\Omega = \{\text{H,T}\}$, where H and T are the events of the coin flipping heads and tails respectively.

Definition 2.2. Given a set Ω, we define the σ-field $\mathcal{F}$ on Ω as a family of subsets of Ω closed under countable union and countable intersection. Equivalently, a σ-field $\mathcal{F}$ on Ω satisfies the following properties:

(1) $\emptyset, \Omega \in \mathcal{F}$.
(2) The complement of a set in $\mathcal{F}$ is also in $\mathcal{F}$. That is, if $\mathcal{A} \in \mathcal{F}$, then $\mathcal{A}^c \in \mathcal{F}$.
(3) The countable union and intersection of sets in $\mathcal{F}$ is also in $\mathcal{F}$. That is, if $\mathcal{A}_1, ..., \mathcal{A}_n, ... \in \mathcal{F}$, then $\bigcap\limits_{i=1}^{\infty} A_i \in \mathcal{F}$ and $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F}$.

In probability, such a field can be understood as the information available to us. It represents the set of events, an event being a set of possible outcomes from an experiment.

Definition 2.3. Given a topological space X, a Borel σ-field $\mathfrak{B}(X)$ is defined as the smallest σ-field containing all open subsets of X.

We need σ-field to describe possible events because it allows us to assign to each set in $\mathcal{F}$ a probability, that is, a value between 0 and 1 (inclusive). We now turn to mathematically defining the notion of probability of an event, which has its commonly understood meaning.

Definition 2.4. A probability measure P on $\mathcal{F}$ is a real-valued function on $\mathcal{F}$ where:

(1) $P(\Omega)= 1$
(2) if $\mathcal{A} \in \mathcal{F}$, then $P(\mathcal{A})= 1 - P(\mathcal{A}^c)$
(3) if $\mathcal{A}_1, ..., \mathcal{A}_n, ... \in \mathcal{F}$ are pairwise disjoint, then $P(\bigcup\limits_{i=1}^{\infty} A_i)= \sum_{n=1}^{\infty} P(\mathcal{A}_i)$

Definition 2.5. The tuple $(\Omega, \mathcal{F}, P)$ is called a probability space or probability triple.

Definition 2.6. Given a probability space $(\Omega, \mathcal{F}, P)$, a function $f : \Omega \to \mathbb{R}$ is said to be $\mathcal{F}$-measurable if, for all Borel sets $U \subset \mathbb{R}$:

$$f^{-1}(U) := \{\omega \in \Omega : f(\omega) \in U\} \in \mathcal{F}$$

Definition 2.7. Given a probability space $(\Omega, \mathcal{F}, P)$, a random variable X is a $\mathcal{F}$-measurable function $X : \Omega \to \mathbb{R}$. To each random variable corresponds a Borel measure in $\mathbb{R}$ defined by $\mu_X(B) = P(X^{-1}(B))$, which we call the distribution of X. We write that $\mathcal{F}_X$, the σ-field generated by X, is the set $\mathcal{F}_X = \{X^{-1}(A) : A \in \mathfrak{B}(\mathbb{R})\}$.

Definition 2.8. Given a real-valued function $f(t)$, we define its quadratic variation $[f, f]$ over the interval $[0, t]$ as the following limit:

$$[f, f] = \lim_{\|P\| \to 0} \sum_{i=1}^{n} (f(t_i) - f(t_{i-1}))^2$$

where P denotes the partition $0 = t_0 < t_1 < ... < t_{n-1} < t_n = t$ of the interval $[0, t]$, and $\|P\|$ denotes the norm of this partition, where $\|P\| = \max_{i \in 1,...,n} (t_i - t_{i-1})$.

Definition 2.9. Similarly, given real-valued functions $f(t)$, $g(t)$, we define their quadratic covariation $[f, g]$ over $[0, t]$ as:

$$[f, g] = \lim_{\|P\| \to 0} \sum_{i=1}^{n} (f(t_i) - f(t_{i-1}))(g(t_i) - g(t_{i-1}))$$

where P denotes the partition $0 = t_0 < t_1 < ... < t_{n-1} < t_n = t$ of the interval $[0, t]$, and $\|P\|$ denotes the norm of this partition, where $\|P\| = \max_{i \in 1,...,n} (t_i - t_{i-1})$.

Definition 2.10. The expectation of a random variable X with respect to a probability measure P is defined as

$$\mathbb{E}(X) = \int_{\Omega} X(\omega)\, dP(\omega) = \int_{\mathbb{R}} x\, d\mu_X(x)$$

Definition 2.11. For a Borel-measurable random variable $X : \Omega \to \mathbb{R}$ and a constant $p \in [1, \infty)$, we define the L^p-norm of X, $\|X\|_p$ as

$$\|X\|_p = \mathbb{E}(|X|^p)^{\frac{1}{p}},$$

and the space corresponding to this norm is defined as:

$$L^p(\Omega) = \{X : \Omega \to \mathbb{R}; \|X\|_p < \infty\}$$

Definition 2.12. We say that a sequence of random variables $\{X_n\}$ converges to X in L^p if $\mathbb{E}(|X_i|^p) < \infty\ \forall i \in \{1, ..., n\}$ and $\mathbb{E}(|X_n - X|^p) \to 0$ as $n \to \infty$

Definition 2.13. A filtration $\mathbb{F}_t$ is a collection of σ-fields $\mathcal{F}_t$ indexed by time t such that, for every $s \leq t$, we have $\mathcal{F}_s \subseteq \mathcal{F}_t$.

Definition 2.14. A stochastic process $\{X_t\}_{t \in T}$ is a set of random variables indexed by time t. If the index t takes on only positive integer values, the stochastic process is said to be discrete. If t can take on any positive real value, the process is continuous. A stochastic process is said to be adapted to a filtration $\mathbb{F}_t$ if, for every t, X_t is $\mathcal{F}_t$ measurable.

In this paper, we will work primarily with continuous stochastic processes.

Definition 2.15. A continuous stochastic process X_t is said to be a martingale with respect to a filtration $\mathbb{F}_t$ if, for all $t \geq 0$ and some $s > 0$, $E(|X_t|) < \infty$, and the following property is true almost surely (with probabiliy 1):

$$\mathbb{E}(X(t + s)|\mathcal{F}_t) = X(t)$$

3. Brownian Motion

Brownian motion was first introduced by British botanist Robert Brown to describe the apparently random motion of pollen in water. As pollen in a cup of water sinks to the bottom, it also moves horizontally, changing directions very frequently. At the molecular level, this movement is determined by the forces of nearly infinitely many and nearly infinitely small water particles. Stock prices can also be modeled by Brownian motion because they are determined by very many buyers and sellers of very small amounts of stock at the same time. We denote Brownian motion by $B(t)$ or, equivalently, by B_t.

3.1. Defining Properties of Brownian Motion.

Let $\{B(t)\}$ be a continuous stochastic process that represents the position of a particle at time t with the following defining properties:

(1) (Independence of increments) for $t > s$, the difference $B(t) - B(s)$ is independent of the past, that is, of all $B(u)$ such that $0 \le u \le s$ or, equivalently, the σ-algebra $\mathcal{F}_s$ generated by $B(u)$, where $u \le s$.
(2) (Normal increments) $B(t) - B(s) \sim N(0, t-s)$. That is, the increment from time s to t is normally distributed with mean 0 and variance $t - s$.
(3) (Path continuity) $B(t)$ is a continuous function with respect to $t \ge 0$.

On the additional condition that $B(0) = 0$, the process is sometimes called a standard Brownian motion or a Wiener process, and it is at times denoted with $W(t)$.

3.2. Some additional Properties of Brownian motion.

Theorem 3.1. *The quadratic variation of Brownian motion* $[B(t), B(t)]$ *over* $[0, t]$ *converges to* t *in* L^2.

Proof. Put the partition that defines the variation as $0 = t_0 < t_1 < ... < t_{n-1} < t_n = t$, and we write $\Delta B_i = B(t_i) - B(t_{i-1})$ to denote the increment from t_{i-1} to t_i. Then we have:

$$\mathbb{E}(([B,B]-t)^2) = \mathbb{E}((\sum_{i=1}^{n}(\Delta B_i)^2 - t)^2)$$

Then, writing t as a telescoping sum $t = \sum_{i=1}^{n}(\Delta t_i)$ using increments $\Delta t_i = t_i - t_{i-1}$, and joining the sums, we obtain:

$$\mathbb{E}((\sum_{i=1}^{n}(\Delta B_i)^2 - t)^2) = \mathbb{E}(\sum_{i=1}^{n}(\Delta B_i^2 - \Delta t_i)^2) \qquad (*)$$

Then, we use the linearity of expectation to bring the expectation inside the sum, we expand the square argument, use linearity again and we obtain the following:

$$\begin{aligned}\mathbb{E}(\sum_{i=1}^{n}(\Delta B_i^2 - \Delta t_i)^2) &= \sum_{i=1}^{n}(\mathbb{E}(\Delta B_i^4 - 2(\Delta B_i)^2(\Delta t_i)^2 + \Delta t_i)^2) \\ &= \sum_{i=1}^{n}(\mathbb{E}(\Delta B_i^4)) - 2\sum_{i=1}^{n}(\mathbb{E}(\Delta B_i)(\Delta t_i)^2) + \sum_{i=1}^{n}(\mathbb{E}(\Delta t_i)^2)\end{aligned}$$

Recall that the increments from times s to t in Brownian motion are normally distributed with mean 0 and variance $t - s$. Then, use the fact that for a normally distributed variable $X \sim N(0, \sigma^2)$, its second moment is $\mathbb{E}(X^2) = \sigma^2$ and its fourth moment is $\mathbb{E}(X^4) = 3\sigma^4$. Also, note that Δt_i is nonrandom and take it out of the expectation by linearity. Then, we get:

$$\mathbb{E}(\sum_{i=1}^{n}(\Delta B_i^2 - \Delta t_i)^2) = \sum_{i=1}^{n}(3(t_i - t_{i-1})^2) - 2\sum_{i=1}^{n}(\Delta t_i)(t_i - t_{i-1}) + \sum_{i=1}^{n}(\Delta t_i)^2$$

We then use the fact that $\Delta t_i = t_i - t_{i-1}$ and simplify the sums to obtain:

$$\mathbb{E}(\sum_{i=1}^{n}(\Delta B_i^2 - \Delta t_i)^2) = 2\sum_{i=1}^{n}(t_i - t_{i-1})^2$$

Now, recall that the quadratic variation is defined by taking the limit of the sum as the partition gets infinitely finer. When $\|P\| \to 0$, each interval in the partition becomes infinitely small, so $t_i - t_{i-1} \to 0 \; \forall i \in 1, ..., n$. Hence, each term in the sum above goes to 0, so the sum goes to 0 and this gives us L^2 convergence as desired:

$$\mathbb{E}(|[B,B] - t|^2) = \mathbb{E}(\sum_{i=1}^{n}(\Delta B_i^2 - \Delta t_i)^2) = \lim_{\Delta t_i \to 0} 2\sum_{i=1}^{n}(t_i - t_{i-1})^2 = 0$$

□

Theorem 3.2. *Brownian motion $B(t)$ is nowhere differentiable with respect to t almost surely.*

A rigorous proof of this fact is beyond the scope of the paper; one such proof can be found in Breiman (1992), p.261-262.

The non-differentiability of Brownian paths is the principal motivation of Itô calculus, which allows us to define an integral with respect to Brownian motion without the fundamental theorem of Riemann calculus. This will be introduced in the next Section.

Theorem 3.3. *Brownian motion $B(t)$ is a martingale with respect to the σ-fields $\mathcal{F}_t$ generated by $\{B_s : s \leq t\}$. That is, $\mathbb{E}(B(t+s)|\mathcal{F}_t) = B(t)$ almost surely for any $s > 0$.*

Proof. Begin with the left-hand side of the equation and add and subtract $B(t)$ in the expectation

$$\mathbb{E}(B(t+s)|\mathcal{F}_t) = \mathbb{E}(B(t) + B(t+s) - B(t)|\mathcal{F}_t)$$

Then, use linearity of conditional expectation to split it in two;

$$\mathbb{E}(B(t) + B(t+s) - B(t)|\mathcal{F}_t) = \mathbb{E}(B(t)|\mathcal{F}_t) + \mathbb{E}(B(t+s) - B(t)|\mathcal{F}_t)$$

Now, we use the fact that increments $B(t+s) - B(s)$ in Brownian motion are independent of $\mathcal{F}_t$ and normally distributed with mean 0. Also, we have $\mathbb{E}(B(t)|\mathcal{F}_t) = \mathbb{E}(B(t))$ in the left summand because $B(t)$ is $\mathcal{F}_t$-measurable, and we are left with:

$$\mathbb{E}(B(t)|\mathcal{F}_t) + \mathbb{E}(B(t+s) - B(t)|\mathcal{F}_t) = B(t) + \mathbb{E}(B(t+s) - B(t)) = B(t) + 0 = B(t)$$

This shows $\mathbb{E}(B(t+s)|\mathcal{F}_t) = B(t)$. □

Theorem 3.4. *Almost surely, Brownian motion $B(t)$ is not monotone on any interval $[a, b] \subset (0, \infty]$.*

Proof. Consider an interval $[a, b] \subset (0, \infty]$ where $a \neq b$. Assume, for the sake of contradiction, that Brownian motion is monotone on this interval. Then, consider a partition $a = t_0 < t_1 < ... < t_{n-1} < t_n = b$ of the interval $[a, b]$. Then, by monotonicity, all increments $B(t_i) - B(t_{i-1})$ must be of the same sign. Since increments are normally distributed, they are either positive with probability 0.5 or negative with probability 0.5. Hence, the probability of B being monotone, that is, all increments having the same sign, is 2^{-n+1}. As we keep taking finer and finer partitions, $n \to \infty$, and this probability goes to 0, since $\lim_{n\to\infty} 2^{-n+1} = 0$. This shows that for any interval $[a, b]$, $B([a, b])$ is not monotone almost surely. We can take the countable union of all intervals with rational endpoints. But then, by the density of $\mathbb{Q}$ in $\mathbb{R}$, we can extend this to all intervals in $(0, \infty]$. □

4. The Itô Integral

Itô's calculus extends many of the methods of calculus to stochastic processes, such as Brownian motion. The Itô integral and Itô's lemma form the basis of Itô calculus, and are described in this Section.

Definition 4.1. Given a function f and a monotone function g that are finite on the interval $[a, b]$ partitioned by $P = a = x_0 < x_1 < ... < x_n = b$ and some intermediate points $\xi_i \in [t_{i-1}, t_i]$, the **Riemann-Stieltjes integral** of f with respect to g over $(a, b]$ is defined as:

$$\int_a^b f(x)dg(x) = \lim_{\|P\|\to 0} \sum_{i=1}^{n} f(\xi_i)(g(t_i) - g(t_{i-1}))$$

Observe that this is a generalization of the Riemann integral, which is just the Stieltjes integral with $g(x) = x$. We will see that the definition of the Itô integral closely resembles that of a Stieltjes integral, but g will be a stochastic process instead.

Remark 4.2. We often understand a Riemann integral as quantifying the "area under the curve" delimited by the integrand $f(x)$. In this interpretation, Riemann sums approximate the area under the curve with rectangles all of which have equal width and variable height, depending on the value of the integrand at that point. At the end we sum up all of these rectangles to get our approximation of the area under the curve. To understand the Stieltjes integral we can think of a similar procedure, except that in the final step, instead of just summing up all the rectangles, we choose to rescale each width of the rectangles by applying some function to them individually. The said function is precisely $g(x)$, which "stretches" or "compresses" the points on the x-axis.

Definition 4.3. We say that a stochastic process $f(t, \omega)$ is elementary, if it is of form $f(t, \omega) = \sum_j e_j(\omega)\mathbb{1}_{[t_j, t_{j+1})}(t)$, where each e_j is a $\mathcal{F}_{t_j}$-measurable random variable and $\mathbb{1}_{[t_j, t_{j+1})}(t)$ is the indicator variable on the interval $[t_j, t_{j+1})$.

4.1. Defining the Itô Integral.

Consider an elementary stochastic process $f(t, \omega)$. Then we can define its Itô integral as

$$\int_S^T f(t, \omega)dB_t(\omega) = \sum_{j\geq 0} e_j(\omega)[B_{t_j+1}, B_{t_j}](\omega)$$

Then, we observe some of the properties of this integral.

4.2. Properties of the Itô Integral.

We can also write the Itô integral as $\int_S^T X(t)dB_t$. Then, it has the following properties:

- (Interval Addition) Given a function $f(t, \omega) \in \mathcal{V}$, and some constant $R \in (S, T) \subset \mathbb{R}$

$$\int_S^T f(t, \omega)dB_t(\omega) = \int_S^R f(t, \omega)dB_t(\omega) + \int_R^T f(t, \omega)dB_t(\omega)$$

- (Linearity) Given functions $f(t, \omega)$, $g(t, \omega) \in \mathcal{V}$, and constants a,b

$$\int_S^T (af(t, \omega) + bg(t, \omega))dB_t = a\int_S^T f(t, \omega)dB_t + b\int_S^T g(t, \omega)dB_t$$

- (Zero Mean Property) Given a function $f(t,\omega) \in \mathcal{V}$
$$\mathbb{E}\left(\int_S^T f(t,\omega)dB_t\right) = 0$$
- (Isometry) Given a function $f(t,\omega) \in \mathcal{V}$
$$\mathbb{E}\left(\int_S^T f(t,\omega)dB_t\right)^2 = \int_S^T \mathbb{E}(f^2(t,\omega))dt$$

Note that the Isometry and Zero Mean properties hold only under the additional assumption that $f(t,\omega)$ is bounded.
Properties 1 and 2 can be proven by simple algebraic manipulations, so we omit their proofs. We instead turn to proving properties 3 and 4.

Proof. (Zero Mean Property)
Consider an elementary function $f(t,\omega) = \sum_j e_j(\omega)\mathbb{1}_{[t_j,t_{j+1})}(t)$. Then, observe that, by the Cauchy-Schwarz inequality:

$$\mathbb{E}(|e_j(\omega)(B_{t_j+1}(\omega) - B_{t_j}(\omega))|) < \sqrt{\mathbb{E}(e_j(\omega)^2)\mathbb{E}((B_{t_j+1}(\omega) - B_{t_j}(\omega))^2)} < \infty$$

In particular, the expression must be bounded because we know that $\mathbb{E}(e_j(\omega)^2) < \infty$, and because the increments in Brownian motions are normally distributed, its second moment is bounded as well. Because this inequality holds, we can apply the triangle inequality to the Itô integral of an elementary function. In particular, we get:

$$\mathbb{E}\left(\left|\int_S^T f(t,\omega)dB_t(\omega)\right|\right) = \mathbb{E}\left(\left|\sum_{j\geq 0} e_j(\omega)\Delta B_{t_j}(\omega)\right|\right)$$
$$\leq \mathbb{E}\left(\sum_{j\geq 0} |e_j(\omega)\Delta B_{t_j}(\omega)|\right) = \sum_{j\geq 0}\mathbb{E}(|e_j(\omega)\Delta B_{t_j}(\omega)|) < \infty$$

This shows that the expectation of the Itô Integral exists.
Now, recall that Brownian motion is a martingale (Theorem 3.3), and that each e_j is $\mathcal{F}_{t_j}$-measurable. Then, we have:

$$\mathbb{E}((e_j(\omega)\Delta B_{t_j}(\omega))|\mathcal{F}_{t_i}) = e_j(\omega)\mathbb{E}(\Delta B_{t_j}(\omega)|\mathcal{F}_{t_i}) = 0$$

this implies that $\mathbb{E}((e_j(\omega)\Delta B_{t_j}(\omega))|) = 0$, which, in turn, implies that each term in the expectation of the Itô integral is 0. Hence, the Itô Integral has mean zero. □

Proof. (Itô Isometry)
We know that each e_j that defines an elementary process $f(t,\omega)$ is L^2, so $f(t,\omega)$ must also be L^2, and we can safely expand the following expectation:

$$\mathbb{E}\left(\int_S^T f(t,\omega)dB_t\right)^2 = \mathbb{E}\left(\sum_{j\geq 0}(e_j(\omega)\Delta B_{t_j}(\omega))\right)^2$$
$$= \sum_{j\geq 0}\mathbb{E}(e_j^2(\omega)(\Delta B_{t_j}(\omega))^2) + 2\sum_{j\neq i}\mathbb{E}(e_j(\omega)e_i(\omega)\Delta B_{t_j}(\omega)\Delta B_{t_i}(\omega))$$

We can again use the martingale property of Brownian motion. Consider the first sum of the previous equation:

$$\sum_{j\geq 0}\mathbb{E}(e_j^2(\omega)(\Delta B_{t_j}(\omega))^2) = \sum_{j\geq 0}\mathbb{E}(\mathbb{E}(e_j^2(\omega)(\Delta B_{t_j}(\omega))^2)|\mathcal{F}_{t_j})$$
$$= \sum_{j\geq 0}\mathbb{E}(e_j^2(\omega)\mathbb{E}((\Delta B_{t_j}(\omega))^2|\mathcal{F}_{t_j}) = \sum_{j\geq 0}\mathbb{E}(e_j^2(\omega)\mathbb{E}(t_i - t_{i-1})|\mathcal{F}_{t_j})$$
$$= \sum_{j\geq 0}\mathbb{E}(e_j^2(\omega)(t_i - t_{i-1})) = \int_S^T \mathbb{E}(f^2(t,\omega)dB_t(\omega))$$

We can do this on the second sum of the expansion of the expectation squared of the Itô integral, and this would give us that each term in the sum equals zero, so that the entire sum equals zero. Hence, we showed: $\mathbb{E}\left(\int_S^T f(t,\omega)dB_t\right)^2 = \int_S^T \mathbb{E}(f^2(t,\omega))dt$.
□

What about continuous functions? It turns out that we can extend the definition of the Itô integral for a wider class of functions.

Consider the set $\mathcal{V}$ of functions $y(t,\omega)$ such that $y(t,\omega)$ is bounded and continuous with respect to t.

Lemma 4.4. *If $y(t,\omega)$ is a function where $y \in \mathcal{V}$, then y can be described as an L^2-limit of a sequence of elementary functions $f_n(t,\omega)$, and consequently all properties of the Itô integral of elementary functions carry over to Itô integrals of any $y \in \mathcal{V}$. See Oksendal (2003), p.27-28 for a proof of this lemma.*

Then, observe that, for any $y(t,\omega) \in \mathcal{V}$, we write its Itô integral as the L^2 limit:

$$\int_S^T y(t,\omega)dB_t(\omega) = \lim_{n\to\infty}\int_S^T f_n(t,\omega)dB_t(\omega)$$

where $f_n(t,\omega)$ is a sequence of elementary functions that converges in L^2 to $y(t,\omega) \in \mathcal{V}$. Also, note that the set of functions for which the integral is defined can be further expanded to the set of functions that are $\mathcal{H}_t$-adapted, where $\mathcal{H}_t$ is a field with respect to which B_t is a martingale (see Oksendal (2003), p.34).

Remark 4.5. Because the Itô integral is defined similarly to a Stieltjes integral, we can interpret the Itô integral to simply be a random version of the Stieltjes integral. However, an Itô integral can also be understood as the profit from trading stock. Suppose $B(t)$ represents the price of stock and $f(t,\omega) = X_{t_i}$ is the amount of stock held in the interval $[t_i, t_{i+1})$. At each time t_i, a trader pays $X_{t_i}B_{t_i}$ for X_{t_i} stocks, and at time t_{i+1}, all X_{t_i} stocks are sold, totaling a net profit of $X_{t_i}(B_{t_{i+1}} - B_{t_i})$. Then, again at time t_{i+1} the trader buys $X_{t_{i+1}}$ stocks and repeats the process until $t = T$. At the end, the total profit equals precisely the value of the Itô integral. This interpretation also shows why we choose the left endpoint e_i instead of any other value at a point in $[t_i, t_{i+1})$. We choose how much stock to buy at time t_i based on the price at the price it has at that time B_{t_i}, and being able to buy it based on any other later price would imply that we are able to see into the future. It's for this reason that Itô calculus is especially useful in finance.

5. Itô's Lemma

Even though we have arrived to the definition of the Itô integral, we still do now know how to compute one. As in the case of Riemann integration, the definition of the integral alone is not useful in practice, and we need to define other tools or techniques to evaluate integrals. Itô's lemma is the tool that allows us to integrate a class of stochastic processes called Itô processes.

Definition 5.1. A stochastic process X_t is called a 1-dimensional Itô process on the probability triple $(\Omega, \mathcal{F}, P)$ if it satisfies the following equation for some u, v:

$$X_t = X_0 + \int_0^t u(s,\omega)ds + \int_0^t v(s,\omega)dB_s$$

or the equivalent equation in differential form:

$$dX_t = u(t,\omega)dt + v(t,\omega)dB_t$$

Here $\int_0^t |u(s,\omega)|ds < \infty$ and $\int_0^t v^2(s,\omega)dB_s < \infty$ almost surely.

5.1. **1-dimensional Itô Lemma.**

Given an Itô process X_t, and a bounded, twice continuously differentiable function $g(t,x) : [0,\infty) \times \mathbb{R} \to \mathbb{R}$ with bounded partial derivatives, then $Y_t = g(t, X_t)$ is an Itô process with:

$$dY_t = \frac{\partial g}{\partial t}(t,X_t)dt + \frac{\partial g}{\partial x}(t,X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)(d{X_t}^2)$$

Proof. Because g's partials are bounded, we can approximate g with a second-degree Taylor polynomial, which gives:

$$g(t,X_t) = g(0,X_0) + \sum_j \frac{\partial g}{\partial t}\Delta t_j + \sum_j \frac{\partial g}{\partial t}\Delta X_j + \sum_j \frac{1}{2}\frac{\partial^2 g}{\partial t^2}(\Delta t_j)^2$$
$$+ \sum_j \frac{1}{2}\frac{\partial^2 g}{\partial t \partial x}(\Delta t_j)(\Delta X_j) + \frac{1}{2}\sum_j \frac{\partial^2 g}{\partial x^2}(\Delta X_j)^2 + \cdots$$

Where all partial derivatives are evaluated at point (t_j, X_j), and $\Delta t_j = t_{j+1} - t_j$, $\Delta X_j = X_{j+1} - X_j$, and the ... denotes the remainder term of form $o(|\Delta t_j|^2 + |\Delta X_j|^2)$ Observe that, as $\Delta t_j \to 0$, we have:

$$\sum_j \frac{\partial g}{\partial t}(t_j, X_j)\Delta t_j \to \int_0^t \frac{\partial g}{\partial t}(s, X_s)ds$$

$$\sum_j \frac{\partial g}{\partial x}(t_j, X_j)\Delta X_j \to \int_0^t \frac{\partial g}{\partial x}(s, X_s)dX_s$$

Then, using the definition of Itô process X_t, we can expand the fifth sum in the Taylor expansion as follows:

$$\sum_j \frac{\partial^2 g}{\partial x^2}(\Delta X_j)^2 = \sum_j \frac{\partial^2 g}{\partial x^2}u_j^2(\Delta t_j)^2 + 2\sum_j \frac{\partial^2 g}{\partial x^2}u_j v_j(\Delta t_j)(\Delta B_j) + \sum_j \frac{\partial^2 g}{\partial x^2}v_j^2(\Delta B_j)^2$$

where u_j and v_j are evaluated at (t_j, ω). Observe that the first sums with terms $(\Delta t_j)^2, (\Delta t_j)(\Delta X_j)$ both become irrelevant in our sum because they go to 0 much more rapidly than the terms with either only Δt_j or ΔX_j in our starting Taylor expansion. It can, however, be shown that this is not true of the third term in the sum with $(\Delta X_j)^2$ We now turn to precisely this term. Put $a_j = a(t_j) = \frac{\partial^2 g}{\partial x^2}(t_j, X_{t_j})v^2(t_j, \omega)$. Then, consider:

$$\mathbb{E}((\sum_j a_j(\Delta B_j)^2 - \sum_j a_j \Delta t_j)^2) = \sum_{i,j} \mathbb{E}((a_i a_j((\Delta B_i)^2 - \Delta t_i)(\Delta B_j)^2 - \Delta t_j))$$

When $i > j$, the factors $(\Delta B_i)^2 - \Delta t_i$,$a_i a_j((\Delta B_j)^2 - (\Delta t_j))$ are independent, and we have:

$$\begin{aligned}\mathbb{E}((a_i a_j((\Delta B_i)^2 - \Delta t_i)(\Delta B_j)^2 - \Delta t_j)) &= \mathbb{E}(a_i a_j((\Delta B_j)^2 - \Delta t_j))\mathbb{E}((\Delta B_i)^2 - \Delta t_i) \\ &= \mathbb{E}(a_i a_j((\Delta B_j)^2 - \Delta t_j))(\Delta t_i - \Delta t_i) = 0\end{aligned}$$

Since the Brownian increment from 0 to t_j is normally distributed with variance t_j. A similar argument can also be applied to the terms where $j > i$. Hence, we only care about the terms where $i = j$, and, as $\Delta t_j \to 0$ our previous equation becomes:

$$\sum_i \mathbb{E}(a_i^2((\Delta B_j)^2 - \Delta t_j)^2) = \sum_j \mathbb{E}(a_j^2)\mathbb{E}((\Delta B_j)^4 + 2(\Delta t_j)(\Delta B_j)^2 + (\Delta t_j)^2)$$

$$= \sum_j \mathbb{E}(a_j^2)(3(\Delta t_j)^2 - 2(\Delta t_j)^2 + (\Delta t_j)^2) = \sum_j \mathbb{E}(a_j^2)(2(\Delta t_j)^2) \to 0$$

This shows that, as $\Delta t_j \to 0$:

$$\sum_i a_i(\Delta B_j)^2 \to \int_0^t a(s)ds \text{ in } L^2(P)$$

Sometimes, we summarize this final step with the shorthand and somewhat misused notation $(dB_t)^2 = dt$. Since we showed that all explicit terms in the original Taylor expansion are finite, we need not worry about the remainder terms. Thus, we have proven Itô's lemma in its entirety. □

Example 5.2. Find $\int_0^t B_s dB_s$.

From regular calculus, we guess that our final answer will have a $\frac{B_t^2}{2}$ term, and we take it from there. Set $X_t = B_t$ and $g(t, x) = \frac{x^2}{2}$. Then $Y_t = g(t, B_t) = \frac{B_t^2}{2}$ and we apply Itô's Lemma and get:

$$dY_t = d(\frac{B_t^2}{2}) = B_t dB_t + \frac{1}{2}dt$$

Integrating both sides gives:

$$\frac{B_t^2}{2} = \int_0^t B_s dB_s + \frac{t}{2} \Rightarrow \int_0^t B_s dB_s = \frac{B_t^2}{2} - \frac{t}{2}$$

5.2. **Itô's Lemma in higher dimensions.** We now formulate Itô's lemma when working with processes in dimensions higher than one.

Let $B(t, \omega) = \langle B_1(t, \omega), ..., B_m(t, \omega)\rangle$ be the m-dimensional Brownian motion. If the processes $X_i = u_i(t, \omega)dt + v_{ij}(t, \omega)dB_j$ are Itô processes for all $1 \le i \le n$ and $1 \le j \le m$, then $X(t) = \langle X_1, ..., X_n\rangle$ is an n-dimensional Itô process. If $g(t, \omega) = \langle g_1(t, \omega), ..., g_p(t, \omega)\rangle$ is a twice differentiable function from $[0, \infty) \times \mathbb{R}^n$ to $\mathbb{R}^p$, then the process $Y(t, \omega) = g(t, X(t))$ is also an Itô process, and for each of its components, we have:

$$dY_k = \frac{\partial g_k}{\partial t}(t,X)dt + \sum_i \frac{\partial g_k}{\partial x_i}(t,X)dX_i + \frac{1}{2}\sum_{i,j}\frac{\partial^2 g_k}{\partial x_i \partial x_j}(t,X)d[X_i,X_j]$$

A proof of Itô's lemma in multiple dimensions is similar to that in 1 dimension and we omit it.

Theorem 5.3. ***(Itô's Integration by Parts)***
Given two Itô processes X_t, Y_t, then the following holds:

$$\int_0^T X_t dY_t = X_T Y_T - X_0 Y_0 - \int_0^t Y_t dX_t - \int_0^T d[X_t, Y_t]$$

Proof. Apply the multi-dimensional Itô formula to the function $g(x,y) = xy$. Then

$$d(X_tY_t) = d(g(X_t,Y_t)) = \frac{\partial g}{\partial x}(X_t,Y_t)dX_t + \frac{\partial g}{\partial y}(X_t,Y_t)dY_t$$

$$+\frac{1}{2}\frac{\partial^2 g}{\partial x^2}(X_t,Y_t)(dX_t)^2 + \frac{1}{2}\frac{\partial^2 g}{\partial y^2}(X_t,Y_t)(dY_t)^2 + \frac{\partial^2 g}{\partial y \partial x}(X_t,Y_t)d[X_t,Y_t]$$

$$= Y_t dX_t + X_t dY_t + dX_t dY_t$$

Integrating both sides, this gives us:

$$X_T Y_T = X_0 Y_0 + \int_0^T Y_t dX_t + \int_0^T X_t dY_t + \int_0^T d[X_t, Y_t]$$

which corresponds to the statement of Integration by parts after rearranging the terms. Note that if X_t, Y_t are Itô processes with $dX_t = adt + bdB_t$, $dY_t = \alpha dt + \beta dB_t$, then $d[X_t, Y_t] = b\beta dt$. □

6. Stochastic Differential Equations

Differential equations are used to describe change in a system. Stochastic differential equations are used when noise is accounted for in the model of a system. In this section, we will define Stochastic differential equations (SDEs), we will consider solutions to SDEs, and we will show that, under certain conditions, solutions exist and are unique.

In this section, we will look at stochastic differential equations of form $\frac{dX_t}{dt} = \mu(t, X_t) + \sigma(t, X_t)B_t$, where $\mu, \sigma : \mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathbb{R}$ respectively denote the drift term and the noise term of the differential equations. The possible solutions X_t to such differential equations are Itô processes.

6.1. Existence and Uniqueness of Solutions.

For the purpose of the next theorem, we will consider a constant $T \in (0, \infty)$ and functions $\mu(t,x) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ $\sigma(t,x) : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ satisfying:

(1) (Linear Growth Condition) $|\mu(t,x)| + |\sigma(t,x)| \leq C(1+|x|)$ for some constant C.
(2) (Locally Lipschitz) $|\mu(x,t) - \mu(y,t)| + |\sigma(x,t) - \sigma(y,t)| < D|x-y|$ for some constant D.

Theorem 6.1. *Let X_0 be a random variable that is independent of $B(t), 0 \le t \le T$ and such that $\mathbb{E}(X_0^2) < \infty$. Then, the stochastic differential equation*

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t \tag{6.2}$$

with initial condition X_0 has a unique solution X_t that is continuous with respect to t and adapted to the filtration $\mathcal{F}_s^Z$, generated by Z and B_s, where $s \le t$.

Lemma 6.3. *(Gronwall's inequality)*
Given nonnegative functions $f(t)$, constants C,A where $A \ge 0$, and if , $\forall t \in [0,T]$, $f(t) \le C + \int_0^t f(s)ds$ holds, then, $\forall t \in [0,T]$:

$$f(t) \le Ce^{At}$$

See Klebner (2005) p. for a proof of this fact.

Proof. We first show that if solutions to (6.2) exist, they must be unique. Consider the solutions $X_1(t,\omega)$, $X_2(t,\omega)$ with initial conditions Z_1, Z_2 respectively. Put $a(s,\omega) = \mu(s, X_1(s,\omega)) - \mu(s, X_2(s,\omega))$ and $b(s,\omega) = \sigma(s, X_1(s,\omega)) - \sigma(s, X_2(s,\omega))$. Then, using the fact that X_1, X_2 are solutions, we have

$$\mathbb{E}(|X_1(s,\omega) - X_2(s,\omega)|^2) = \mathbb{E}\left[\left(Z_1 - Z_2 + \int_0^t a ds + \int_0^t b dB_s\right)^2\right]$$

Then, using the Cauchy-Schwarz inequality for sums $\sum_i u_i v_i < (\sum_i u_i^2)(\sum_i v_i^2)$ where, for $i \in \{1,2,3\}$ $u_i = 1$, and $v_1 = Z_1 - Z_2$, $v_2 = \int_0^t a ds$, $v_3 = \int_0^t b dB_s$, we have:

$$\mathbb{E}(|X_1(s,\omega) - X_2(s,\omega)|^2) \le 3\mathbb{E}(|Z_1 - Z_2|^2) + 3\mathbb{E}\left[\left(\int_0^t a ds\right)^2\right] + 3\mathbb{E}\left[\left(\int_0^t b dB_s\right)^2\right]$$

Then, using the Cauchy-Schwarz inequality for integrals $\int_c^d fg = \int_c^d f^2 \int_c^d g^2$ on the first integral, and Itô isometry on the second integral, we get:

$$\mathbb{E}(|X_1(s,\omega) - X_2(s,\omega)|^2) \le 3\mathbb{E}(|Z_1 - Z_2|^2) + 3t\mathbb{E}\left[\int_0^t a^2 ds\right] + 3\mathbb{E}\left[\int_0^t b^2 dB_s\right]$$

Then, observe that by assumption 2. (Lipschitz continuity), we have $a^2 = (\mu(t, X_1(t,\omega)) - \mu(t, X_2(t,\omega)))^2 \le D^2(X_1 - X_2)^2$. An analogous reasoning can be used on b^2, and it gives us:

$$\mathbb{E}(|X_1(s,\omega) - X_2(s,\omega)|^2) \le 3\mathbb{E}(|Z_1 - Z_2|^2) + 3(1+t)D^2 \int_0^t \mathbb{E}((X_1(s,\omega) - X_2(s,\omega))^2)ds$$

Set $v(s) = \mathbb{E}(|X_1(s,\omega) - X_2(s,\omega)|^2)$, $C = 3\mathbb{E}(|Z_1 - Z_2|^2)$, $A = 3(1+T)D^2$, and apply Gronwall's inequality to get:

$$\text{if } v(t) \le C + A\int_0^t v(s)ds, \text{ then } v(t) \le Ce^{At}$$

However, since X_1, X_2 are both solutions to equation (6.2), their initial conditions must be the same, so $Z_1 = Z_2$. But then, $Ce^{At} = 0$, so $v(t) = 0$ as well, which implies $\mathbb{E}((X_1 - X_2)^2) = 0$. Therefore, $P(|X_1 - X_2|) = 0$ for all rational $t \in [0,T]$. However, because the map $t \mapsto |X_1 - X_2|$ is continuous, it follows that $P(|X_1 - X_2|) = 0$ can be extended to all values of $t \in [0,T]$, and this terminates the proof of uniqueness of the solution.
We now turn to showing the existence of a solution. Let $Y_t^{(0)} = X_0$, and define

inductively $Y_t^{(k+1)} = X_0 + \int_0^t \mu(s, Y_s^{(k)})ds + \int_0^t \sigma(s, Y_s^{(k)})dB_s$. Then, using a similar reasoning as that in our proof of the uniqueness of a solution, we have $\mathbb{E}((Y_t^{(k+1)} - Y_t^{(k)})^2) \leq (1+T)3D^2 \int_0^t \mathbb{E}((Y_s^{(k)} - Y_s^{(k-1)})^2)ds$ for $k \geq 1$, $t \leq T$. Then consider $\mathbb{E}((Y_t^{(1)} - Y_t^{(0)})^2)$. By definition, we have:

$$\begin{aligned}\mathbb{E}((Y_t^{(1)} - Y_t^{(0)})^2) &= \mathbb{E}\left(\left(\int_0^t \mu(s, X_0)ds + \int_0^t \sigma(y, X_0)dB_s\right)^2\right)\\ &\leq \mathbb{E}\left(\left(\int_0^t \mu(s, X_0)ds\right)^2\right) + \mathbb{E}\left(\left(\int_0^t \sigma(y, X_0)dB_s\right)^2\right)\end{aligned}$$

Then, using the linear growth condition and Itô isometry:

$$\begin{aligned}\mathbb{E}((Y_t^{(1)} - Y_t^{(0)})^2) &\leq \mathbb{E}\left(\left(\int_0^t C(1+|X_0|)ds\right)^2\right) + \mathbb{E}\left(\int_0^t C^2(1+|X_0|)^2\right)\\ &\leq 2C^2(t+t^2)(1+\mathbb{E}(|X_0|^2)) \leq A_1 t\end{aligned}$$

where A_1 is a constant depending on C,T, $\mathbb{E}(|X_0|^2)$. By induction, we can then show that $\mathbb{E}((Y_t^{(k+1)} - Y_t^{(k)})^2) \leq \frac{A_2^{k+1}t^{k+1}}{(k+1)!}$ for positive integers n and nonnegative k, where A_2 is a constant depending on C,T,D, $\mathbb{E}(|X_0|^2)$. Then, consider the Lebesgue measure λ on $[0, T]$ and $m > n \geq 0$, and we have, using telescoping sums and the triangle inequality:

$$\|Y_t^m - Y_t^n\|_{L^2(\lambda\times P)} = \left\|\sum_{k=n}^{m-1} Y_t^{k+1} - Y_t^k\right\|_{L^2(\lambda\times P)} \leq \sum_{k=n}^{m-1} \|Y_t^{k+1} - Y_t^k\|_{L^2(\lambda\times P)}$$

Then, using the definition of L^2 norm, we have:

$$\begin{aligned}\sum_{k=n}^{m-1} \|Y_t^{k+1} - Y_t^k\|_{L^2(\lambda\times P)} &= \sum_{k=n}^{m-1}\left(\mathbb{E}\left(\int_0^t (Y_s^{(k+1)} - Y_s^k)^2 ds\right)\right)^{\frac{1}{2}}\\ &\leq \sum_{k=n}^{m-1}\left(\int_0^t \frac{A_2^{k+1}s^{k+1}}{(k+1)!}ds\right)^{\frac{1}{2}} = \sum_{k=n}^{m-1}\left(\frac{A_2^{k+1}t^{k+2}}{(k+2)!}\right)^{\frac{1}{2}} \to 0\end{aligned}$$

since each term in the sum goes to 0 as $m, n \to \infty$. This shows that $Y_t^{(n)}$ converges in $L^2(\lambda \times P)$. We will now show that the limit of this sequence is a solution. Put $X_t = \lim_{n\to\infty} Y_t^{(n)}$; note that X_t is $\mathcal{F}_t^Z$-measurable because each $Y_t^{(n)}$ is also $\mathcal{F}_t^Z$-measurable. Consider the following:

$$\mathbb{E}\left(\left(\int_0^t |\mu(s, X_s) - \mu(s, Y_s^{(n)})|ds\right)^2\right) \leq t^{\frac{1}{2}}\mathbb{E}\left(\int_0^t \mathbb{E}(|\mu(s, X_s) - \mu(s, Y_s^{(n)})|^2)ds\right)^{\frac{1}{2}}$$

By the Lipschitz condition of $\mu(t, x)$, we then have, for some constant D

$$t^{\frac{1}{2}}\mathbb{E}\left(\int_0^t \mathbb{E}(|\mu(s, X_s) - \mu(s, Y_s^{(n)})|^2)ds\right)^{\frac{1}{2}} \leq D^2\mathbb{E}(|X_s - Y_s^{(n)}|^2) \to 0$$

since $\mathbb{E}(|X_s - Y_s^{(n)}|^2) \to 0$ as $n \to \infty$. This shows that:

$$\int_0^t \mu(s, Y_s^{(n)})ds \to \int_0^t \mu(s, X_s)ds \text{ in } L^2$$

We can use a similar argument, combined with Itô isometry, to show that:

$$\int_0^t \sigma(s, Y_s^{(n)})dB_s \to \int_0^t \sigma(s, X_s)dB_s$$

This shows that $\forall t \in [0,t]$, $X_t = X_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s$, so X_t is indeed a solution to our starting equation (6.2). □

6.2. **The Black–Scholes Model.**

We now look at one of the most famous applications of Stochastic calculus in finance: the Black–Scholes Model. The Black–Scholes Model is used to describe the theoretical price of a European-style call option. A call option is a contract that gives a buyer the right —but not the obligation— to buy an asset (stock, commodity, or index) at a predetermined price, the strike price, until an agreed-upon time, the expiration date. The model assumes that there is no arbitrage in the market, meaning that there is no way for an agent to make profit without risk. We also assume that an agent's portfolio is self-financing, i.e. that an agent starts trading with an initial amount of money, and that the agent cannot receive additional money. The only way to get more money is by shorting an asset, i.e. selling it before having it and owing it at some point in the future.

Begin with a stock of price S_t, which follows the stochastic equation $dS_t = \mu S_t dt + \sigma S_t dB_t$, and suppose we have a call option with expiration date $T > 0$ and strike price K. Then, assume that there are no transaction fees and the payoff from the call option is $V_T = \max(S_T - K, 0)$, and assume that $V_t = v(t, S_t)$ for some function v and $t < T$. Assume that there are also bonds available on the market, whose value Y_t follows the equation $dY_t = rY_t dt$ for some interest rate $r > 0$. Denote our portfolio by O_t, where $O_t = X_t S_t + Y_t$, where X_t denotes the amount of stock we hold. Then, by the self-financing assumption, we have that $dO_t = X_t dS_t + rY_t dt$. Then, we want $dO_t - dv(t, S_t) = 0$. Use Itô's lemma and the fact that $O_t = Y_t + X_t S_t = v(t, S_t)$ to observe:

$$dO_t - dv(t, S_t) = \mu X_t S_t dt + \sigma X_t S_t dB_t + rY_t dt - v'(t, S_t)dS_t - \frac{1}{2}v''(t, S_t)\sigma^2 S_t^2 dt - \dot{v}(t, S_t)dt$$

$$dO_t - dv(t, S_t) = \mu X_t S_t dt + \sigma X_t S_t dB_t + r(v(t, S_t) - X_t S_t)dt - v'(t, S_t)(\mu S_t dt + \sigma S_t dB_t) - \frac{1}{2}v''(t, S_t)\sigma^2 S_t^2 dt - \dot{v}(t, S_t)dt$$

Here $\dot{v}$ is the time derivative of v and v', v'' are the first and second order derivatives of v with respect to S_t. Then, choose $X_t = v'(t, S_t)$ and, since we assumed the no arbitrage condition in the market, observe that all terms in the expression with dt equal 0. Then, we have

$$rv(t, S_t) - rS_t v'(t, S_t) - \frac{1}{2}v''(t, S_t)\sigma^2 S_t^2 - \dot{v}(t, S_t) = 0$$

Rearranging and multiplying by -1, and letting $S_t = x$ this gives us precisely the Black–Scholes formula, with the boundary condition $v(T, s) = \max(S_t, 0)$:

$$\dot{v}(t, x) + rxv'(t, x) + \frac{1}{2}\sigma^2 x^2 v''(t, x) - rv(t, x) = 0$$

To solve this kind of partial stochastic differential equation, many methods may be used, such as the Feymann-Kac formula, or numerical methods like the finite difference method. The interested reader should see Lawler (2006), p.221 for a discussion of these methods.

Acknowledgments

I am deeply thankful to my mentor, Jinwoo Sung, for his help, kindness, and patience. I am grateful to Professor Lawler for his lectures in probability and analysis and to Professor May for organizing the REU. A special thanks goes to Shreya helping me with LATEX.

References

[1] Oksendal, B., (2003). Stochastic Differential Equations: An Introduction with Applications. Springer.
[2] Klebner, F., (2005). Introduction to Stochastic Calculus with Applications. Monash University.
[3] Lawler, G., (2023). Stochastic Calculus: An Introduction with Applications. https://www.math.uchicago.edu/~lawler/finbook.pdf. Accessed 18 August 2023.
[4] Lawler, G., (2006). Introduction to Stochastic Processes. Chapman & Hall/CRC.
[5] Breiman, L., (1992). Probability. SIAM.