

An Introduction to Stochastic Differential Equations

Adair Antonio da Silva Neto

June 29, 2022

Contents

1	Introduction	1
2	Mathematical Preliminaries	2
2.1	Probability Spaces	2
2.2	Distribution and Density of a Random Variable	4
2.3	Expected Value and Variance	5
2.4	Stochastic Processes	8
2.5	Brownian Motion	9
2.6	Lebesgue Integral	10
3	Ito Integrals	12
3.1	Construction	12
3.2	Properties	18
3.3	Extensions	18
3.4	Stratonovich integral	19
4	The Ito Formula and the Martingale Representation Theorem	21
5	Stochastic Differential Equations	22
	References	23

1 Introduction

Present the kind of problems we're going to work on, and sketch the way.

2 Mathematical Preliminaries

2.1 Probability Spaces

A probability space starts with a set Ω called the **sample space**, which is, intuitively, a list of all possible outcomes of an experiment. Each $\omega \in \Omega$ is a **sample point**, and each subset $A \subseteq \Omega$ is an **event**.

To filter ‘well-behaved’ subsets, where it will be possible to measure a probability, the following definition is necessary.

Definition 2.1 (σ -algebra). If Ω is a set, then a σ -**algebra** \mathfrak{F} on Ω is a family of subsets of Ω satisfying:

1. $\emptyset \in \mathfrak{F}$.
2. If $A \in \mathfrak{F}$, then $A^c \in \mathfrak{F}$.
3. If $A_1, A_2, \dots \in \mathfrak{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{F}$.

The pair (Ω, \mathfrak{F}) is said to be a **measurable space**, and any subset $B \subseteq \Omega$ that also belongs to \mathfrak{F} is called a **measurable set**.

Example 2.1 (Examples of σ -algebras). The following are σ -algebras.

1. The family $\{\emptyset, \Omega\}$ is called the **trivial σ -algebra**, and is the smallest one possible.
2. The power set $\mathfrak{P}(\Omega)$ is called the **discrete σ -algebra**, and is the largest one possible.
3. The family $\{\emptyset, \Omega, A, A^c\}$ is the **σ -algebra generated by the set A** and is usually denoted by \mathfrak{F}_A .

But the most important σ -algebra for probability theory is the Borel σ -algebra, which will be denoted by \mathfrak{B} . Taking $\Omega = \mathbb{R}$, the Borel σ -algebra is generated by the intersection of all σ -algebras containing the real line intervals.

Notice that the Borel σ -algebra contains all open sets, closed sets, and all their countable operations with union \cup , intersection \cap , and their complements c . This is the smallest σ -algebra containing all open subsets.

More generally, if \mathcal{U} is the collection of all open subsets of a topological space Ω , then \mathfrak{B} is called the **Borel σ -algebra** on Ω . The elements $B \in \mathfrak{B}$ are called **Borel sets**.

Given a measurable space, it's possible to assign each outcome to a probability.

Definition 2.2 (Probability Measure). Let (Ω, \mathfrak{F}) be a measurable space. A **probability measure** P is the function

$$P : \mathfrak{F} \longrightarrow [0, 1]$$

satisfying

1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
2. (σ -additivity). If $A_1, A_2, \dots \in \mathfrak{F}$ and $A_i \cap A_j = \emptyset$, for $i \neq j$ (i.e. are mutually exclusive), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

And the triple $(\Omega, \mathfrak{F}, P)$ is called a **probability space**.

In order to restrict the functions to sets in the σ -algebra, the following definition will be needed.

Definition 2.3 (\mathfrak{F} -measurable function). Given a probability space $(\Omega, \mathfrak{F}, P)$, a function Y from the sample space Ω to \mathbf{R}^n is called \mathfrak{F} -measurable if

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

for all open sets $U \in \mathbf{R}^n$ (i.e. for all Borel sets $U \subseteq \mathbf{R}^n$). In other words, the inverse image of U is in the σ -algebra.

And to attach numerical values to each $\omega \in \Omega$, define the following function.

Definition 2.4 (Random variable). An \mathfrak{F} -measurable function $X : \Omega \longrightarrow \mathbf{R}^n$ is called a **random variable** on a complete probability space $(\Omega, \mathfrak{F}, P)$.

This definition means that if we know which event U in the \mathfrak{F} has occurred, then we know which value of X has occurred.

Consider the measurable space $(\Omega, \mathfrak{P}(\Omega))$, and let X be a random variable with values $x_i, i = 1, 2, \dots, k$. The sets

$$A_i = \{\omega \in \Omega : X(\omega) = x_i\} \subseteq \Omega$$

form a partition of Ω , and the σ -algebra generated by this partition is called the **σ -algebra generated by X** .

Notice that this is the smallest σ -algebra that contains all the sets A_i . This σ -algebra is often denoted by \mathfrak{F}_X . Intuitively, this σ -algebra represents all information available about the sample point ω by observing X .

2.2 Distribution and Density of a Random Variable

However, given only the probability measure P , it is not always immediate how to compute the probability of a given interval, set or value. For that purpose, the following three definitions will be useful.

Definition 2.5 (Distribution). The **distribution** of a random variable X is a function $\mu_X : \Omega \rightarrow \mathbf{R}^n$ defined as

$$\mu_X(B) = P[X^{-1}(B)] = P[X \in B] = P(\{\omega : X(\omega) \in B\})$$

Definition 2.6 (Distribution Function). The **cumulative distribution function** (CDF) of a random variable X is defined as

$$F_X(x) = P[X \leq x] = P(\{\omega : X(\omega) \leq x\})$$

where $x \in \mathbf{R}$. Notice that F is non-decreasing, and right-continuous, and approaches 0 at $-\infty$ and 1 at $+\infty$.

Definition 2.7 (Joint Distribution). If X and Y are random variables, then their **joint distribution function** is

$$F_{X,Y}(x,y) = P[X \leq x, Y \leq y]$$

where $x, y \in \mathbf{R}$.

The distribution function gives the probability that the random variable X is on the interval $(-\infty, x]$. This also allows to look for an interval $[a, b]$ using $F_X(b) - F_X(a)$.

While the distribution of X returns the probability of an event $\{\omega : X(\omega) \in B\}$.

Definition 2.8 (Density function). A random variable X has a **probability density function** (PDF) $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ if f is a measurable function and

$$F_X(x) = P[X \leq x] = \int_{-\infty}^x f(y)dy$$

2.3 Expected Value and Variance

Another useful tool for any given random variable is to know its mean value and how much it varies. This motivates the following.

Definition 2.9 (Expectation). Let $(\Omega, \mathfrak{F}, P)$ be a probability space and X a random variable. If $\int_{\Omega} |X(\omega)| dP(\omega) < \infty$, then the **expected value** (or **mean value**) of X with respect to P is

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbf{R}^n} x d\mu_X(x)$$

Notice that the expectation is linear. I.e., if X and Y are integrable and a and b are constants, then

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

If f is a Borel measurable function and $\int_{\Omega} |f(X)| dP(\omega) < \infty$, then

$$\mathbb{E}[f(X)] = \int_{\Omega} f(X(\omega)) dP(\omega) = \int_{\mathbf{R}^n} f(x) d\mu_X(x)$$

Theorem 2.1 (Chebychev's inequality).

$$P[|X| \geq \lambda] \leq \frac{1}{\lambda^p} \mathbb{E}[|X|^p]$$

for all $\lambda \geq 0$.

Definition 2.10 (Variance). Let $\mu := \mathbb{E}[X]$, i.e., the expected value of a random variable X . Then the **variance** of X is given by

$$\text{Var}[X] = \mathbb{E}[|X - \mu|^2] = \int_{\Omega} |X - \mu|^2 dP(\omega)$$

Definition 2.11 (Covariance). Let X and Y be integral random variables. Let $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$. If XY is integrable, then the **covariance** of X and Y is:

$$\text{Cov}[X, Y] = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mu_X \mu_Y$$

Notice that $\text{Var}[X] = \text{Cov}[X, X]$.

The variance of X can be computed more simply using the following theorem.

Theorem 2.2. Let X be a random variable. Then,

1. $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
2. $\text{Var}[aX + b] = a^2 \text{Var}[X]$.
3. $\text{Cov}[X, Y] = \text{Cov}[Y, X]$.
4. $\text{Cov}[aX + bY, Z] = a \text{Cov}[X, Z] + b \text{Cov}[Y, Z]$.
5. $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2 \text{Cov}[X, Y]$.

Definition 2.12 (Independence). Two events $A, B \in \mathfrak{F}$ are **independent** if

$$P[A \cap B] = P[A] \cdot P[B]$$

More generally, any collection of events A_i , $i = 1, 2, \dots$, is called independent if for any $n \in \mathbb{N}$ and any choice of indices i_k , $k = 1, 2, \dots$,

$$P\left(\bigcap_{k=1}^n A_{i_k}\right) = \prod_{k=1}^n P(A_{i_k})$$

Definition 2.13 (Conditional Expectation). Let X and Y be random variables. Then, given that $f_Y(y) > 0$, the **conditional distribution** of X given $Y = y$ is given by the conditional density

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

And the **conditional expectation** of X given $Y = y$ is

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x f(x|y) dx$$

The intuition behind this definition is to build an estimate of the random variable X given the information available in Y .

Some important and very useful properties of conditional expectation are listed below.

Theorem 2.3. Let $X, Y : \Omega \longrightarrow \mathbf{R}^n$ be random variables with $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|Y|] < \infty$, $\mathfrak{H} \subseteq \mathfrak{F}$ a σ -algebra, and let $a, b \in \mathbf{R}$.

1. $\mathbb{E}[aX + bY | \mathfrak{H}] = a\mathbb{E}[X | \mathfrak{H}] + b\mathbb{E}[Y | \mathfrak{H}]$.
2. $\mathbb{E}[\mathbb{E}[X | \mathfrak{H}]] = \mathbb{E}[X]$.
3. $\mathbb{E}[X | \mathfrak{H}] = X$ if X is \mathfrak{H} -measurable.
4. $\mathbb{E}[X | \mathfrak{H}] = \mathbb{E}[X]$ if X is independent of \mathfrak{H} .
5. $\mathbb{E}[Y \cdot X | \mathfrak{H}] = Y \cdot \mathbb{E}[X | \mathfrak{H}]$ if Y is \mathfrak{H} -measurable.
6. If $\mathfrak{G} \subseteq \mathfrak{H}$ is a σ -algebra, then

$$\mathbb{E}[X | \mathfrak{G}] = \mathbb{E}[\mathbb{E}[X | \mathfrak{H}] | \mathfrak{G}]$$

7. If $\varphi : \mathbf{R} \longrightarrow \mathbf{R}$ is convex and $\mathbb{E}[|\varphi(X)|] < \infty$, then

$$\varphi(\mathbb{E}[X | \mathfrak{H}]) \leq \mathbb{E}[\varphi(X) | \mathfrak{H}]$$

Definition 2.14 (Infinitely often). If $A_1, A_2, \dots, A_n, \dots$ are events in the probability space, then the event

$$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \{\omega \in \Omega : \omega \text{ belongs to infinitely many of the } A_n\}$$

is called A_n **infinitely often**, or simply ' A_n **i.o.**'.

The next lemma helps us check if some sequence of events occurs infinitely often.

Lemma 2.4 (Borel-Cantelli). If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then

$$P(A_n \text{ i.o.}) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) = 0$$

The most important distribution in this text is the **normal** (or **gaussian**) distribution.

Definition 2.15 (Normal Distribution). If random variable X has mean μ , variance σ^2 and a density function of the form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$$

then X is said to have a **normal distribution**, which we denote by $X \sim N(\mu, \sigma^2)$.

Notice that X can be normalized (i.e. transformed into a distribution of the form $N(0, 1)$) by taking $Z = \frac{X-\mu}{\sigma}$.

Definition 2.16 (Lipschitz and Hölder Conditions). A function f is said to be **Hölder continuous** of order α , $0 < \alpha \leq 1$, on an interval $[a, b] \subseteq \mathbf{R}$ if there exists a constant $K > 0$, such that for all $x, y \in [a, b]$ we have

$$|f(x) - f(y)| \leq K|x - y|^\alpha$$

If $\alpha = 1$, then the function f is said to be **Lipschitz continuous**.

2.4 Stochastic Processes

Definition 2.17 (Stochastic Process). A **stochastic process** is a collection of random variables $\{X(t)\}$ parametrized by time $t \in T$.

For each point $\omega \in \Omega$, the mapping $t \longrightarrow X_t(\omega)$ is the respective **sample path**, also called **realization** or **trajectory**.

Definition 2.18 (Filtration). A filtration \mathbf{F} is a collection of σ -algebras

$$\mathbf{F} = \{\mathfrak{F}_0, \mathfrak{F}_1, \dots, \mathfrak{F}_t, \dots, \mathfrak{F}_T\}$$

such that $\mathfrak{F}_t \subseteq \mathfrak{F}_{t+1}$.

The idea behind this definition is to model the flow of information. As the time t passes, the observer has more information, and does not lose any previous data. Notice that this implies finer partitions of the sample space Ω .

Given a stochastic process $\{X(t)\}$, let \mathfrak{F}_t be the σ -algebra generated by the random variables X_s , $s = 0, \dots, t$. Since $\mathfrak{F}_t \subseteq \mathfrak{F}_{t+1}$, these σ -algebras form a filtration called the **natural filtration** of the process $\{X(t)\}$, and contain all available information of the process up to the time t .

Definition 2.19 (Martingale). A stochastic process $\{M_t\}$ on $(\Omega, \mathfrak{F}, P)$ is called a **martingale** with respect to a filtration \mathfrak{M}_t if

1. M_t is \mathfrak{M}_t -measurable for all t .
2. $\mathbb{E}[|M_t|] < \infty$ for all t .
3. $\mathbb{E}[M_s | \mathfrak{M}_t] = M_t$ for all $s \geq t$.

Intuitively, a martingale is a stochastic process in which the future has no tendency to go up or down. In other words, the expected value of any time in the future is equal to the value of the process at the present time.

Theorem 2.5 (Doob's martingale inequality). If M_t is a martingale such that the mapping $t \rightarrow M_t(\omega)$ is continuous a.s., then for all $p \geq 1$, $T \geq 0$ and $\lambda > 0$,

$$\mathbb{P}\left[\sup_{0 \leq t \leq T} |M_t| \geq \lambda\right] \leq \frac{1}{\lambda^p} \mathbb{E}[|M_T|^p]$$

2.5 Brownian Motion

In 1828 the Scottish botanist Robert Brown described the irregular motion of pollen grains suspended in fluid, that is now known as the Brownian Motion. To describe this mathematically, we use a stochastic process $B_t(\omega)$, which is the position at time t of the particle ω .

This can also be understood as the model for the cumulative effect of 'noise'. And is also called Wiener process, after N. Wiener, who formalized mathematically the Brownian motion.

Definition 2.20 (Brownian Motion). A Brownian motion $\{B(t)\}$ is a stochastic process with the following properties:

1. (Independence of increments). $B(t) - B(s)$, for $t > s$, are independent. That means that this difference does not depend on the past.
2. (Normal increments). $B(t) - B(s)$ has normal distribution with mean 0 and variance $t-s$. Notice that by taking $s = 0$ we have $B(t) - B(0) \sim N(0, t)$.
3. (Continuity of paths). $B(t)$, $t \geq 0$, are continuous functions of t .

The existence and continuity of Brownian motion is proved using Kolmogorov's extension and continuity theorem [Oks13].

Theorem 2.6 (Quadratic variation). If $t \geq s$,

$$\mathbb{E}[(B_t - B_s)^2] = t - s$$

More generally (in \mathbb{R}^n):

$$\mathbb{E}[(B_t - B_s)^2] = n(t - s)$$

Theorem 2.7.

$$\mathbb{E}[f(B_t)] = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} f(x) e^{-\frac{x^2}{2t}} dx$$

Theorem 2.8. The expected value of odd moments of B_t (other than one) are zero. And the even moments are given by

$$\mathbb{E}[B_t^{2k}] = \frac{(2k)!}{2^k k!} t^k$$

Last but not least, two important identities, useful for further calculations are:

$$\begin{aligned} B_j(B_{j+1} - B_j) &= \frac{1}{2}[B_{j+1}^2 - B_j^2 - (B_{j+1} - B_j)^2] \\ B_j^2(B_{j+1} - B_j) &= \frac{1}{3}(B_{j+1}^3 - B_j^3) - B_j(B_{j+1} - B_j)^2 - \frac{1}{3}(B_{j+1} - B_j)^3 \end{aligned}$$

2.6 Lebesgue Integral

Definition 2.21. A space $L^p(\Omega, \mathcal{F}, P)$ is the space of measurable functions in which the p -th power of the absolute value is Lebesgue integrable. I.e., functions of the form

$$\|f\|_p := \left(\int_{\Omega} |f|^p d\mu \right)^{1/p} < \infty$$

We can also write this as

$$\|X\|_p := \sqrt[p]{\mathbb{E}[|X|^p]} < \infty$$

A sequence of random variable can converge in different ways. The following definitions are given in an increasing order of strength.

Definition 2.22 (Convergence in Distribution). A sequence of random variables $\{X_n\}$ **converges in distribution** to X if their distribution functions $F_{X_n}(x)$ converge to the distribution function $F_X(x)$ at any point of continuity of F_X .

In other words, for any bounded function $f : \mathbf{R} \longrightarrow \mathbf{R}$ we have

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$$

I.e.,

$$\lim_{n \rightarrow \infty} \int_{\Omega_n} f(X_n) dP_n = \int_{\Omega} f(X) dP$$

Definition 2.23 (Convergence in Probability). A sequence of random variables $\{X_n\}$ **converges in probability** to X if for any $\varepsilon > 0$, we have

$$\lim_{n \rightarrow \infty} P[|X_n(\omega) - X(\omega)| > \varepsilon] = 0$$

Definition 2.24 (Convergence Almost Surely). A sequence of random variables $\{X_n\}$ **converges almost surely (a.s.)** to X if for any ω outside a set of probability zero we have that

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

Alternatively, we can write

$$P[X_n(\omega) \rightarrow X(\omega)] = 1$$

Definition 2.25 (L^p -Convergence). A sequence of random variables $\{X_n\}$ **converges in L^p** to X if $\{X_n\} \subseteq L^p$, $p \in [1, \infty)$, and

$$\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0$$

I.e., $\mathbb{E}[|X_n|^p] < \infty$ and

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

3 Ito Integrals

In this section, we'll present how the Itô Integral is constructed, showing the intuition behind it and the necessary steps for the proof. However, for full proofs, please refer to [Oks13] or [S⁺04].

3.1 Construction

We are interested in solving differential equations of the form

$$\frac{dX}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot \text{'noise'}$$

This noise will be represented by some stochastic process W_t . Our goal is to satisfy the following properties:

- If $t_1 \neq t_2$, then W_{t_1} and W_{t_2} are independent.
- The collection $\{W_t\}$ is stationary, i.e., the joint distribution of any collections of W_t does not depend on t .
- $\mathbb{E}[W_t] = 0$ for all t .

However, a process W_t satisfying the first two conditions cannot have continuous paths. So our first goal is to replace W_t with some convenient stochastic process.

Let $0 = t_0 < t_1 < \dots < t_m = t$ and consider the discrete version

$$X_{k+1} - X_k = b(t_k, X_k) \Delta t_k + \sigma(t_k, X_k) W_k \Delta t_k$$

Now, replace $W_k \Delta t_k = \Delta V_k = V_{k+1} - V_k$, where V_t is a stochastic process. Since we want V_t to have stationary independent increments with mean zero, the only suitable process with continuous paths is the Brownian motion. Hence, we take $V_t = B_t$, thus obtaining

$$X_k = X_0 + \sum_{j=0}^{k-1} b(t_j, X_j) \Delta t_j + \sum_{j=0}^{k-1} \sigma(t_j, X_j) \Delta B_j$$

And taking the limit as Δt_j goes to zero, we have

$$X_k = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s$$

The question now is to understand the second integral, which can be generalized as

$$\int_S^T f(t, \omega) dB_t(\omega)$$

where $f : [0, \infty] \times \Omega \longrightarrow \mathbf{R}$.

The idea is to first study a simple class of functions and then extended it to more general functions by taking the limit of the integral of simple functions.

So assume that f is of the form

$$\varphi(t, \omega) = \sum_{j \geq 0} e_j(\omega) \chi_{[j \cdot 2^{-n}, (j+1) \cdot 2^{-n}]}(t)$$

where χ is the characteristic (indicator) function and n is a natural number. Then we can define

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

where

$$t_k = \begin{cases} S & \text{if } k \cdot 2^{-n} < S \\ T & \text{if } k \cdot 2^{-n} > T \\ k \cdot 2^{-n} & \text{if } S \leq k \cdot 2^{-n} \leq T \end{cases}$$

This definition, however, also leads to difficulties, as a consequence of the variations of the paths of B_t .

In order to deal with that difficulty, notice that a given function $f(t, \omega)$ can be approximated by

$$\sum_j f(t_j^*, \omega) \chi_{[t_j, t_{j+1})}(t)$$

where $t_j^* \in [t_j, t_{j+1}]$.

With that, we can define

$$\int_S^T f(t, \omega) dB_t(\omega) = \lim_{n \rightarrow \infty} \sum_j f(t_j^*, \omega) [B_{t_{j+1}} - B_{t_j}](\omega)$$

If we take $t_j^* = t_j$, the left end point, then we obtain the **Ito integral**, denoted by

$$\int_S^T f(t, \omega) dB_t(\omega)$$

If, however, we take $t_j^* = (t_j + t_{j+1})/2$, the mid point, then we obtain the **Stratonovich integral**, denoted by

$$\int_S^T f(t, \omega) \circ dB_t(\omega)$$

One of the advantages of Ito integral is its application in Finance. Intuitively, by taking the leftmost point, one does not have to know the future, like whether a stock goes up or down.

Related to this intuition, and developing the Ito integral, we must restrict ourselves to functions f that only depends on the behaviour of $B_s(\omega)$ up to the time t_j .

Definition 3.1. Let $B_t(\omega)$ be a Brownian motion. We define \mathfrak{F}_t as the σ -algebra generated by the random variables B_s , where $s \leq t$. I.e., \mathfrak{F}_t is the smallest σ -algebra containing all sets

$$\{\omega : B_{t_1}(\omega) \in F_1, \dots, B_{t_k}(\omega) \in F_k\}$$

where $t_j \leq t$ and $F_j \subseteq \mathbf{R}^n$ are Borel sets, $j \leq k = 1, 2, \dots$

The intuition behind this definition is that \mathfrak{F}_t can be thought as the ‘history of B_s up to the time t ’. A function $h(\omega)$ is \mathfrak{F}_t -measurable if h depends only on the values from B_0 up to B_t . I.e., it does not depend on the ‘future’.

Definition 3.2 (\mathfrak{N}_t -Adapted). Let $\{\mathfrak{N}_t\}_{t \geq 0}$ be an increasing family of σ -algebras of subsets of Ω . A process $g(t, \omega) : [0, \infty) \times \Omega \longrightarrow \mathbf{R}^n$ is \mathfrak{N}_t -**adapted** if for each $t \geq 0$ the function $\omega \longrightarrow g(t, \omega)$ is \mathfrak{N}_t -measurable.

Example 3.1. The process $h_1(t, \omega) = B_{t/2}(\omega)$ is \mathfrak{F}_t -measurable and, hence, \mathfrak{F}_t -adapted. Notice that this process only depends on previous information.

However, the process $h_2(t, \omega) = B_{2t}(\omega)$ is not \mathfrak{F}_t -measurable and, hence, it is not \mathfrak{F}_t -adapted. That’s because the process h_2 depends on future information.

Now we can define the class of functions for which it is possible to compute the Ito integral.

Definition 3.3. Let $\mathfrak{V}(S, T)$ be the class of functions

$$f(t, \omega) : [0, \infty) \times \Omega \longrightarrow \mathbf{R}$$

such that

1. The function $(t, \omega) \longrightarrow f(t, \omega)$ is $\mathfrak{B} \times \mathfrak{F}$ -measurable, where \mathfrak{B} denotes the Borel σ -algebra on $[0, \infty)$.
2. The function $f(t, \omega)$ is \mathfrak{F}_t -adapted.
3. $\mathbb{E} \left[\int_S^T f(t, \omega)^2 dt \right] < \infty$.

Given functions $f \in \mathfrak{V}$, in order to define the Ito integral

$$I[f](\omega) = \int_S^T f(t, \omega) dB_t(\omega)$$

we are going to define $I[\varphi]$ for a simple class of functions φ and then show how each $f \in \mathfrak{V}$ can be approximated by φ .

A function $\varphi \in \mathfrak{V}$ is **elementary** if it has the form

$$\varphi(t, \omega) = \sum_j e_j(\omega) \chi_{[t_j, t_{j+1})}(t)$$

Since $\varphi \in \mathfrak{V}$, each function e_j is \mathfrak{F}_j -measurable. And, as we did before, we define

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

To further our development of the Ito integral, we need the following result.

Theorem 3.1 (Ito isometry). If $\varphi(t, \omega)$ is bounded and elementary, then

$$\mathbb{E} \left[\left(\int_S^T \varphi(t, \omega) dB_t(\omega) \right)^2 \right] = \mathbb{E} \left[\int_S^T \varphi(t, \omega)^2 dt \right]$$

Using this isometry, it is possible to extended the previous definition to functions in \mathfrak{V} by the following steps.

1. Let $g(\omega) \in \mathfrak{V}$ be bounded and continuous for each ω . Then there exists elementary functions $\varphi_n \in \mathfrak{V}$ such that

$$\mathbb{E} \left[\int_S^T (g - \varphi_n)^2 dt \right] \longrightarrow 0$$

as $n \rightarrow \infty$.

2. Let $h \in \mathfrak{V}$ be bounded, then there exists bounded functions $g_n \in \mathfrak{V}$ such that $g_n(\omega)$ are continuous for all ω and n , and

$$\mathbb{E} \left[\int_S^T (h - g_n)^2 dt \right] \longrightarrow 0$$

3. Let $f \in \mathfrak{V}$. Then there exists a sequence of functions $\{h_n\} \subseteq \mathfrak{V}$ such that h_n is bounded for each n and

$$\mathbb{E} \left[\int_S^T (f - h_n)^2 dt \right] \longrightarrow 0$$

as $n \rightarrow \infty$.

Please note that we started from bounded and continuous functions, generalized into bounded functions and then generalized further into any function in our class \mathfrak{V} .

With these steps, we define the Ito integral as follows.

Definition 3.4 (The Ito integral). Let $f \in \mathfrak{V}(S, T)$, then the **Ito integral** of f , from S to T is

$$\int_S^T f(t, \omega) dB_t(\omega) = \lim_{n \rightarrow \infty} \int_S^T \varphi_n(t, \omega) dB_t(\omega)$$

where the limit is in $L^2(P)$ and $\{\varphi_n\}$ is a sequence of elementary functions such that

$$\mathbb{E} \left[\int_S^T (f(t, \omega) - \varphi_n(t, \omega))^2 dt \right] \longrightarrow 0$$

as $n \rightarrow \infty$.

This new definition induces a generalized form of the Ito isometry.

Theorem 3.2 (Ito isometry). For all $f \in \mathfrak{V}(S, T)$,

$$\mathbb{E} \left[\left(\int_S^T f(t, \omega) dB_t(\omega) \right)^2 \right] = \mathbb{E} \left[\int_S^T f(t, \omega)^2 dt \right]$$

Corollary 3.3. If $f(t, \omega) \in \mathfrak{V}(S, T)$ and $f_n(t, \omega) \in \mathfrak{V}(S, T)$ and

$$\mathbb{E} \left[\left(\int_S^T f_n(t, \omega) - f(t, \omega) dt \right)^2 \right] \longrightarrow 0$$

as $n \rightarrow \infty$, then

$$\int_S^T f_n(t, \omega) dB_t(\omega) \longrightarrow \int_S^T f(t, \omega) dB_t(\omega)$$

in $L^2(P)$ as $n \rightarrow \infty$.

Example 3.2. We wish to compute the integral

$$\int_0^T B(t) dB(t)$$

Step 1. Let

$$X_n(t) = \sum_{i=0}^{n-1} B_n(t_i^n) [B_n(t_{i+1}^n) - B_n(t_i^n)]$$

be a sequence of elementary functions. Notice that $X_n(t)$ converges to $B(t)$ almost surely as $\max_i(t_{i+1}^n - t_i^n) \rightarrow 0$ by the continuity of $B(t)$.

Step 2. Now notice that

$$\begin{aligned} B_n(t_i) [B_n(t_{i+1}) - B_n(t_i)] &= B_n(t_{i+1}) B_n(t_i) - B_n^2(t_i) + B_n^2(t_{i+1}) - B_n^2(t_{i+1}) \\ &= \frac{1}{2} [B_n^2(t_{i+1}) - B_n^2(t_i) - (B_n(t_{i+1}) - B_n(t_i))^2] \end{aligned}$$

Step 3. With these two results, we can write the original integral as

$$\int_0^T X_n(t) dB(t) = \frac{1}{2} \sum_{i=0}^{n-1} [B_n^2(t_{i+1}) - B_n^2(t_i)] - \frac{1}{2} \sum_{i=0}^{n-1} [B_n(t_{i+1}) - B_n(t_i)]^2$$

Step 4. Since the first sum is a telescopic sum and the second one converges in probability to T by the quadratic variation of Brownian motion, we have

$$\int_0^T X_n(t) dB(t) = \frac{1}{2} B^2(t) - \frac{1}{2} B^2(0) - \frac{1}{2} T = \frac{1}{2} B^2(t) - \frac{1}{2} T$$

Hence, the integral converges in $L^2(P)$ to

$$\int_0^T B(t) dB(t) = \lim_{n \rightarrow \infty} \int_0^T X_n(t) dB(t) = \frac{1}{2} B^2(t) - \frac{1}{2} T$$

3.2 Properties

Before heading to more theoretical and important results, let us notice some natural facts of the Itô integral.

Theorem 3.4. Let $f, g \in \mathfrak{V}(0, T)$ and let $0 \leq S < U < T$. Then

1. $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$ for a.a. ω .
2. $\int_S^T (cf + g) dB_t = c \int_S^T f dB_t + \int_S^T g dB_t$ where $c \in \mathbb{R}$.
3. $\mathbb{E} \left[\int_S^T f dB_t \right] = 0$.
4. $\int_S^T f dB_t$ is \mathfrak{F}_T -measurable.

With the Doob's martingale inequality, it can be proved that the Itô integral can be chosen to depend continuously on t . A proof of this fact is on the third chapter of [Oks13].

Now an important result is that the Itô integral is a martingale.

Theorem 3.5. Let $f(t, \omega) \in \mathfrak{V}(0, T)$ for all T . Then

$$M_t(\omega) = \int_0^t f(s, \omega) dB_s$$

is a martingale w.r.t. \mathfrak{F}_t and for $\lambda, T > 0$

$$\mathbb{P} \left[\sup_{0 \leq t \leq T} |M_t| \geq \lambda \right] \leq \frac{1}{\lambda^2} \mathbb{E} \left[\int_0^T f(s, \omega)^2 ds \right]$$

Summarizing, the Itô integral is continuous, adapted, linear, a martingale, satisfies the Itô isometry and has Quadratic variation.

3.3 Extensions

Using the concept of martingales, we can generalize the Itô integral for a larger class of functions than \mathfrak{V} .

Considering the Definition 3.3, we can relax the condition 2. into

2.' There exists an increasing family of σ -algebras \mathfrak{H}_t such that B_t is a martingale w.r.t. \mathfrak{H}_t and f_t is \mathfrak{H}_t -adapted.

The idea here is that B_t must remain a martingale with respect to the history of f_s .

Another way of extending the Itô integral definition is by weakening the condition 3. into

$$3.' \quad \mathbb{P}\left[\int_S^T f(s, \omega)^2 ds < \infty\right] = 1$$

To understand why this works, let $\mathfrak{W}_{\mathfrak{H}}(S, T)$ denote the class of processes satisfying conditions 1, 2' and 3' above. Then, in the 1-dimensional Brownian motion, for all t there exists step functions $f_n \in \mathfrak{W}_{\mathfrak{H}}$ such that

$$\int_0^t |f_n - f|^2 ds \longrightarrow 0$$

in probability.

Now, since $\int_0^t f_n(s, \omega) dB_s(\omega)$ converges in probability to a random variable and the limit depends only on f , we define

$$\int_0^t f(s, \omega) dB_s(\omega) = \lim_{n \rightarrow \infty} \int_0^t f_n(s, \omega) dB_s(\omega)$$

where the limit is in probability and $f \in \mathfrak{W}_{\mathfrak{H}}$.

3.4 Stratonovich integral

As we saw, the Itô integral is one possible interpretation of the integral

$$\int_S^T f(t, \omega) dB_t(\omega)$$

The Stratonovich integral is another possibility and, in general, leads to different results (except when the integrating function has a derivative). In some cases, the Stratonovich definition may be adequate.

In other cases, Itô's feature of 'not looking in the future' (as [Oks13] puts it), justifies its use in biology and finance, for example. Moreover, Itô integral is a martingale, and Stratonovich's is not. This gives an important computational advantage to Itô's definition.

And, to quote [S⁺04],

However, it [Stratonovich's integral] is inappropriate for finance. In finance, the integrand represents a position in an asset and the integrator represents the price of that asset. We cannot decide at 1:00 p.m. which position we took at 9:00 a.m. We must decide the position at the beginning of each time interval, and the Ito integral is the limit of the gain achieved by that kind of trading as the time between trades approaches zero.

4 The Ito Formula and the Martingale Representation Theorem

5 Stochastic Differential Equations

References

- [Dur19] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.
- [Eva12] Lawrence C Evans. *An introduction to stochastic differential equations*, volume 82. American Mathematical Soc., 2012.
- [Kle12] Fima C Klebaner. *Introduction to stochastic calculus with applications*. World Scientific Publishing Company, 2012.
- [Oks13] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013. 9, 12, 18, 19
- [S⁺04] Steven E Shreve et al. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer, 2004. 12, 19