Notes on Brownian motion and stochastic processes

Ana E. de Orellana

Contents

Most of this material is based on [KS91], [Gar88], and [Eva12].

1. Stochastic processes

A stochastic process is a model of a random phenomenon that depends on time. The randomness is captured by the introduction of a measurable space $(\Omega, \mathscr{F})$ called the *sample space*, on which probability measures can be defined.

Formally, a *stochastic process* is a collection of random variables $X = \{X_t : 0 \leqslant t < \infty\}$ that take values on a second measurable space called the state space $(S, \mathscr{S})$ (which usually is $S = \mathbb{R}^d$, $\mathscr{S} = \mathscr{B}(\mathbb{R}^d)$). For each $t \in [0, \infty)$ (or $t \in [0, T]$ for some time T), $X_t : \Omega \to S$. Fixing $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is the *sample path* of the process X associated with omega.

AEdO was financially supported by the University of St Andrews.

Example 1.1. *Take the height of people as our random variable. In this case Ω is the population and S is $\mathbb{R}_+$. We can measure heights at different times, so fixing one person $\omega \in \Omega$, $X_t(\omega)$ is that person's height for all the times $t \geqslant 0$. That is a stochastic process.*

It is a model of a random experiment whose outcome can be observed continuously in time.

Intuition. *There's nothing 'random' in the definition of a random variable. A random variable is a measurable function that goes from Ω to S, that we don't know how it works, all we know are the outputs. So we can 'see' how this function works by taking the measure of preimages $P(\{\omega : X(\omega) = a\})$, we don't know which ω are being mapped to a, but we can learn their proportion. So a stochastic process a collection of random variables that depend on both ω and t. We usually fix ω and study $X_t(\omega)$ as a function of t.*

1.1. **Filtrations.** We equip our sample space $(\Omega, \mathscr{F})$ with a *filtration*. That is, a non-decreasing family $\{\mathscr{F}_t : t \geqslant 0\}$ of sub σ-Algebras of $\mathscr{F}$. $\mathscr{F}_s \subseteq \mathscr{F}_t \subseteq \mathscr{F}$ for $0 \leqslant s < t < \infty$. We define $\mathscr{F}_\infty = \sigma\big(\cup_{t\geqslant 0} \mathscr{F}_t\big)$[1].

The simplest choice of a filtration is the one generated by the stochastic process itself. Defined as

$$\mathscr{F}_t^X = \sigma(X_s : 0 \leqslant s \leqslant t),$$

is the smallest σ-algebra with respect to X_s is measurable for every $s \in [0, t]$.

We say that the stochastic process X is *adapted* to the filtration $\{\mathscr{F}_t\}$ if, for each $t \geqslant 0$, X_t is an $\mathscr{F}_t$-measurable random variable.

Intuition. *Filtrations are a way to encode the information contained in the history of a stochastic process. If a process is adapted, then all information about the process up to a certain time is contained in the corresponding filtration.*

1.2. **Martingales.** The conditional probability of some event A given another event B, $P(A|B)$, can be thought of changing the probability space to $\widetilde{\Omega} = B$, $\widetilde{\mathscr{F}} = \{C \cap B : C \in \mathscr{F}\}$ and the probability to $\widetilde{P} = P/P(B)$, so that $\widetilde{P}(\widetilde{\Omega}) = 1$. Thus, if we want to calculate the expected value of a variable X given the event B, we should set

$$\mathbb{E}(X|B) = \frac{1}{P(B)} \int_B X \, dP.$$

And what about the expected value of X given another random variable Y? In other words, if 'chance' selects a sample point $\omega \in \Omega$ and all we know about ω is the value $Y(\omega)$, what is our best guess as to the value of $X(\omega)$?

The conditional expectation of X given Y is a $\mathscr{F}^Y$-measurable random variable Z such that

$$\int_A X \, dP = \int_A Z \, dP, \qquad \forall A \in \mathscr{F}^Y.$$

We denote Z by $\mathbb{E}(X|Y)$.

Intuition. *We can understand $\mathbb{E}(X|Y)$ as the information available in the σ-algebra generated by Y, $\mathscr{F}^Y$ (which is the smallest σ-algebra that contains all the information of Y) that we want to use to estimate the values of X.*

[1]When X is a set, we denote by $\sigma(X)$ to the σ-algebra generated by X. When X is a random variable, $\sigma(X)$ is the sigma-algebra generated by the random variable, that is $\sigma(X) = \sigma\big(\{X^{-1}(A) : A \in \mathscr{F}\}\big)$

A discrete-time *martingale* is a discrete-time stochastic process such that for any time t, $\mathbb{E}|X_t| < \infty$ and $\mathbb{E}(X_{t+1}|X_1, X_2, \dots, X_t) = X_t$. That is, the conditional expected value of the next observation given all the past observations is equal to the most recent observation. Where the conditional expectation of X given A is

$$\mathbb{E}(X|A) = \sum_x xP(X = x|A) = \sum_x x\frac{P(\{X = x\} \cap A)}{P(A)}.$$

Consider a real-valued stochastic process X on a probability space $(\Omega, \mathscr{F}, P)$ adapted to a filtration $\{\mathscr{F}_t\}$ and such that $\mathbb{E}|X_t| < \infty$ for any $t \geqslant 0$. The process $\{X_t, \mathscr{F}_t : 0 \leqslant t < \infty\}$ is a sub(super)martingale if for every $0 \leqslant s < t < \infty$, that P-a.e. $\mathbb{E}(X_t|\mathscr{F}_s) \geqslant X_s$ ($\leqslant$). We say that $\{X_t, \mathscr{F}_t : 0 \leqslant t < \infty\}$ is a martingale if it's both a sub and super martingale.

Example 1.2. *(From Wikipedia) Consider a gambler who wins £1 when a coin copes up heads and loses £1 when the coin comes up tails. Suppose that heads comes up with probability p.*

- *If $p = 1/2$, the gambler on average neither wins nor loses money, and the gambler's fortune over time is a martingale.*
- *If $p < 1/2$, the gambler loses on average, so his fortune over time is a supermartingale.*
- *If $p > 1/2$, the gambler wins on average, so his fortune over time is a submartingale.*

1.3. **Stopping times.** Consider a measurable space $(\Omega, \mathscr{F})$ with a filtration $\{\mathscr{F}_t\}$. A random variable T is a *stopping time* of the filtration, if the event $\{\omega : T(\omega) < t\} \in \mathscr{F}_t$ for every $t \geqslant 0$.

Intuition. *The occurrence of the event $T = t$ only depends on the values of $X_1, \dots, X_t$. The time a gambler leaves the table depends only on the past winnings and loses, not on the future ones.*

1.4. **Sationary processes with independent increments.** A stochastic process X_t is said to be *strictly stationary* if its finite-dimensional distributions are invariant under time displacements. That is, for any $0 \leqslant t < \infty$ such that $t_j, t_j + t$ are in $[0, \infty)$ for all j,

$$F_{t_1+t,\dots,t_n+t}(x_1, \dots, x_n) = F_{t_1,\dots,t_n}(x_1, \dots, x_n).$$

If also $X_t \in L^2$, then for each $s, t \in [0, \infty)$, $E(X_t) = \mu$ and $\mathrm{Cov}(X_t, X_s) = C(t - s)$.

Another type of process are those with *independent increments*, also called *additive process*, which mean that, for any finite sequence $\{t_j\} \subset [0, \infty)$ with $t_j < t_{j+1}$, the differences $X_{t_{j+1}} - X_{t_j}$ are independent. In these cases the probability is determined by the distribution of X_t and $X_t - X_s$ for $t > s$.

2. Brownian motion

A standard one-dimensional *Brownian motion* is a continuous, adapted process $W = \{W_t, \mathscr{F}_t : 0 \leqslant t < \infty\}$, defined on some probability space $(\Omega, \mathscr{F}, P)$ with the properties that $W_0 = 0$ almost surely and for any two pairs of times $0 \leqslant s < t$, the increment $W_t - W_s$ is independent of $\mathscr{F}_s$[2], and is normally distributed with mean zero and variance $t - s$. The filtration $\{\mathscr{F}_t\}$ is not necessarily the one induced by the stochastic process W. In fact, for some applications (such as stochastic differential equations), a larger filtration is needed. It's an example of a strictly stationary process with independent increments (see Section 1.4)

For each $n \in \mathbb{N}$ the Brownian motion satisfies

$$\mathbb{E}|W_t - W_s|^{2n} \approx_n C_n|t - s|^n. \tag{2.1}$$

This is useful because of the following continuity theorem by Kolmogorov.

[2]That means that if $0 \leqslant s_1 < t_1 \leqslant s_2 < t_2$ then $W_{t_1} - W_{s_1}$ is independent of $W_{t_2} - W_{s_2}$

Theorem 2.1. *Suppose that a process $X = \{X_t : 0 \leqslant t \leqslant T\}$ on a probability space $(\Omega, \mathscr{F}, P)$ satisfies the condition*

$$\mathbb{E}|X_t - X_s|^\alpha \lesssim |t-s|^{1+\beta}, \quad 0 \leqslant s,t \leqslant T,$$

for some $\alpha, \beta > 0$. Then there exists a continuous modification $\widetilde{X} = \{\widetilde{X}_t : 0 \leqslant t \leqslant T\}$ of X, which is locally Hölder continuous with exponent γ for every $\gamma \in (0, \beta/\alpha)$, i.e.

$$P\Big(w : \sup_{0<t-s<h(w)} \frac{|\widetilde{X}_t(w) - \widetilde{X}_s(w)|}{|t-s|^\gamma} \leqslant \delta\Big) = 1,$$

where $h(w)$ is an a.s. positive random variable and $\delta > 0$ is some constant.

Given different times $t_1, \dots, t_n$ we are interested in knowing the probability that a sample path of Brownian motion takes values between a_i and b_i for each time t_i. That is, what is

$$P(a_1 \leqslant W_{t_1} \leqslant b_1, \dots, a_n \leqslant W_{t_n} \leqslant b_n)?$$

The following theorem will give an answer to a more general problem.

Theorem 2.2. *Let W_t be the standard Brownian motion. Then for all $n \in \mathbb{N}$, all choices of times $0 = t_0 < t_1 < \cdots < t_n$ and each function $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\mathbb{E}\big(f(W_{t_1}, \dots, W_{t_n})\big) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{x_1^2}{2t_1}} \cdots \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}} e^{-\frac{(x_n - x_{n-1})^2}{2(t_n - t_{n-1})}} \, dx_n \cdots dx_1.$$

This comes from the fact that

$$P(a \leqslant W_t \leqslant b) = \frac{1}{\sqrt{2\pi t}} \int_a^b e^{-\frac{x^2}{2t}} \, dx.$$

Setting $f(x_1, \dots, x_n) = 1_{[a_1,b_1]}(x_1) \cdots 1_{[a_n,b_n]}$ we get

$$P(a_1 \leqslant W_{t_1} \leqslant b_1, \dots, a_n \leqslant W_{t_n} \leqslant b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \frac{1}{\sqrt{2\pi t_1}} e^{-\frac{x_1^2}{2t_1}} \cdots \frac{1}{\sqrt{2\pi(t_n - t_{n-1})}} e^{-\frac{(x_n - x_{n-1})^2}{2(t_n - t_{n-1})}} \, dx_n \cdots dx_1.$$

2.1. **Markov processes.** Let X_t be a stochastic process with state space $\mathbb{R}^d$ and times $0 \leqslant t < \infty$. For $t_1 \leqslant t_2$, define

$$\mathscr{F}([t_1, t_2]) = \sigma(X_t : t_1 \leqslant t \leqslant t_2).$$

That is, $\mathscr{F}([t_1, t_2])$ is the σ-algebra that encodes all the information between the times t_1 and t_2.

The process X_t is called a *Markov process* if for $t_0 \leqslant s \leqslant t \leqslant T$ and all $B \in \mathscr{B}$,

$$P(X_t \in B | \mathscr{F}([t_0, s])) = P(X_t \in B | X_s),$$

with probability 1.

Intuition. *The Markov property says that the probable future state of the system at any time $t > s$ is independent of the past behaviour of the system at times $t < s$, given the present state at time s.*

Intuition. *The process only 'knows' its value at time s and does not 'remember' how it got there.*

The following definition will establish some notation that we will use from now on.

Definition 2.3 (Probability kernel)**.** *A measure kernel from a measurable space $(\mathbb{X}, \mathscr{F}_\mathbb{X})$ to another measurable space $(\mathbb{Y}, \mathscr{F}_\mathbb{Y})$ is a function $P : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}^+$ such that*

- *For any $Y \in \mathbb{Y}$, $P(x,Y)$ is $\mathscr{F}_{\mathbb{X}}$ measurable.*
- *For any $x \in \mathbb{X}$, $P(x,Y)$ is a measure on $(\mathbb{Y}, \mathscr{F}_{\mathbb{Y}})$.*

We will write the integral of a function $f : \mathbb{Y} \to \mathbb{R}$ with respect to this measure as

$$\int f(Y)P(x,dY) := \int f(Y)P(x,Y)\,dY.$$

A well-known equation by Chapman–Kolmogorov says that for fixed s,t,u, with $s \leqslant u \leqslant t$,

$$P(s,x,t,B) = \int_{\mathbb{R}^d} P(u,y,t,B)P(s,x,u,dy).$$

where $P(s,X_s,t,B) = P(X_t \in B|X_s)$ is the *transition probability* between the times $s < t$. The Chapman–Kolmogorov equation indicates that the one-step transition probability can be written in terms of all possible combinations of two-step transition probabilities with respect to any arbitrary intermediate time.

Brownian motion is a Markov process with stationary transition probability

$$P(t,x,B) = \int_B (e\pi t)^{-\frac{d}{2}} e^{-\frac{|y-x|^2}{2t}}\,dy.$$

And

$$\mathbb{E}|W_t - W_s|^4 = 3(t-s)^2,$$

this shows by 2.1 that W has a version with continuous sample paths. We'll see a bit more of this in the next subsection.

2.2. **Modulus of continuity.** A function g is called a *modulus of continuity* for the function f if $g(\delta) \to 0$ as $\delta \to 0$ and $|t-s| \leqslant \delta$ imply $|f(t)-f(s)| \leqslant g(\delta)$ for all sufficiently small δ. Because of the Hölder condition that we know Brownian motion satisfies (see (2.1) and Theorem 2.1), its modulus of continuity cannot be any larger than a constant multiple of δ^γ for any $\gamma \in (0,1/2)$. Levy proved that with

$$g(\delta) = \sqrt{2\delta \log(1/\delta)}, \quad \delta > 0,$$

$cg(\delta)$ is a modulus of continuity for almost every Brownian path on $[0,1]$ if $c > 1$, but is a modulus for almost no Brownian path on $[0,1]$ if $0 < c < 1$.

3. Diffusion processes

Brownian motion has the property that the distribution of $X_t - X_s$ given the σ-algebra $\mathscr{F}_s$ of information up to time s, is Gaussian with mean 0 and variance $t-s$. One can visualise a Markov process X_s for which the corresponding conditional distribution of the increment is approximately Gaussian with mean $ha(s,X_s)$ and variance $hB(s,X_s)$. In such case, $a(s,x)$ would be a d-dimensional vector and $B(s,x)$ would be a symmetric positive semi-definite matrix. $hB(s,X_s)$ would then be approximately the conditional variance matrix of the vector $X_t - X_s$.

$$\lim_{t\to s} \frac{1}{t-s} \int_{|y-x|>\varepsilon} P(s,x,t,dy) = 0,$$

$$\lim_{t\to s} \frac{1}{t-s} \int_{|y-x|\leqslant\varepsilon} (y-x)P(s,x,t,dy) = a(s,x),$$

$$\lim_{t\to s} \frac{1}{t-s} \int_{|y-x|\leqslant\varepsilon} (y-x)(y-x)^{\mathrm{T}}P(s,x,t,dy) = B(s,x).$$

The functions $a(s, X_s)$ and $B(s, X_s)$ are called the coefficients of the diffusion process. In particular, a is referred to as the *drift vector* and B the *diffusion matrix*.

The transition probability $P(s, x, t, B)$ of a diffusion process is uniquely determined by the drift and the diffusion coefficients of the process (under some regularity conditions).

4. Stochastic integration

We will replace ordinary differential equations of the form $\frac{dx}{dt} = f(t, x)$ with a random differential equation

$$\frac{dX}{dt} = F(t, X, Y), \tag{4.1}$$

where $Y = Y_t$ represents some stochastic input process explicitly. A solution to (4.1) is an indexed family of functions depending on time. If the sample path structure of Y is sufficiently pathological (e.g. not integrable), then (4.1) must be reinterpreted, that is, we cannot interpret (4.1) as an ordinary differential equation along each path.

For example, a solution of the equation

$$\frac{dX}{dt} = f(t, X) + g(t, X)\mathcal{N},$$

where $\mathcal{N}$ is a Gaussian white noise process, should be the solution of

$$X_t = X_{t_0} + \int_{t_0}^{t} f(s, X_s)\, ds + \int_{t_0}^{t} g(s, X_s)\mathcal{N}_s\, ds.$$

However, the last integral cannot be defined given the irregularity of $\mathcal{N}$.

To deal with this problem, the last integral will be replaced with

$$\int_{t_0}^{t} g(s, X_s)\, dW_s, \tag{4.2}$$

where W_s is the Brownian motion. This is motivated by the fact that (formally) $\frac{dW_t}{dt} = \mathcal{N}_t$ (white noise is the time derivative of Brownian motion). So now we need only to worry about how to interpret (4.2).

Intuition (Why **white** noise?). *If X_t is a stochastic process with $\mathbb{E}(X_t^2) < \infty$ for all $t \geqslant 0$, we define the autocorrelation function of X_t by*

$$r(t, s) := \mathbb{E}(X_t X_s).$$

If $r(t, s) = c(t - s)$ for some function $c : \mathbb{R} \to \mathbb{R}$, then X_t is stationary. A white noise process is a Gaussian, stationary process, with $c(t) = \delta_0(t)$. In general, we define the Fourier transform of the autocorrelation function,

$$f(\xi) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\xi t} c(t)\, dt,$$

to be the spectral density of the process.

For white noise,

$$f(\xi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\xi t} \delta_0(t)\, dt = \frac{1}{2\pi} \quad \forall \xi.$$

Thus, the spectral density of the white noise $\mathcal{N}$ is flat, that is, all frequencies contribute equally in the correlation function. Just as all colours contribute equally to make white light.

Recall that to define the Riemann-Stieltjes integration we take increments with respect to a function F, $F(t_j) - F(t_{j-1})$ and define

$$\int_a^b g(t)\, dF(t) \approx \sum_{j=1}^{n} g(t_{j-1})F(t_j) - F(t_{j-1}).$$

The condition that we need on F in the Riemann-Stieltjes construction is for it to have bounded variation. Any differentiable function with continuous derivative $f(t) = F'(t)$ has finite variation, and $\int_a^b g(t)\, dF(t) = \int_a^b g(t)f(t)\, dt$, so we write $dF(t) = f(t)\, dt$.

We would hope to do the same for integrals of the form $\int_a^b g(t)\, dW_t$ by using approximations $\sum_{i=1}^n g(t_i)(W_{t_i} - W_{t_{i-1}})$, where each $g(t_i)$ is a random variable. There are two difficulties with this. The first one is that the random variable $g(t_i)$ may not be measurable with respect to the σ-algebra generated by $W_{t_{i-1}}$, and the convergence of the sum is dependent on the choice of points t_i. The second problem is clear from the following proposition.

Proposition 4.1. *The paths of Brownian motion are a.s. not of bounded variation.*

So defining the integral as we know using classic calculus is quite difficult. We will study the solution of Itô, which is choosing t_i to be the left extremes of the intervals.

4.1. **Itô calculus.** Let $\{W_t : 0 \leqslant t \leqslant T < \infty\}$ be the standard Brownian motion process. We know that $W_t - W_s \sim N(0, t-s)$, then $\mathbb{E}(W_t - W_s)^2 = \mathbb{V}(W_t - W_s) = t - s$. Therefore, if $0 = t_0 < t_1 < \cdots < t_n = T$,

$$\mathbb{E}\Big(\sum_{k=1}^{n} W_{t_k} - W_{t_{k-1}}\Big)^2 = \sum_{k=1}^{\infty}(t_k - t_{k-1}) = T. \tag{4.3}$$

And so, since

$$\int_0^T W_s\, dW_s \approx \sum_{k=1}^{n} W_{t_{k-1}}(W_{t_k} - W_{t_{k-1}}) = \frac{1}{2}W_t^2 - \frac{1}{2}\sum_{k=1}^{n}(W_{t_k} - W_{t_{k-1}})^2. \tag{4.4}$$

Letting $n \to \infty$, the last term of (4.4) converges to T from (4.3). Therefore,

$$\int_0^T W_s\, dW_s(s) = \frac{1}{2}W_t^2 - \frac{1}{2}T$$

Intuition. *We are saying that $(dW_t)^2 = dt$, because the definition of Itô's integral involves convergence in the quadratic sense.*

Theorem 4.2. *The integral $\int_a^b W_t\, dW_t$ interpreted in the Itô sense satisfies*

$$\mathbb{E}\Big(\int_a^b W_t\, dW_t\Big) = 0$$
$$\mathbb{E}\Big|\int_a^b W_t\, dW_t\Big|^2 = \tfrac{1}{2}(b^2 - a^2).$$

What differences Itô's calculus from others is the choice of $h(t_k) = W_{t_k}$. If we had chosen instead the midpoint $h(t_k) = B(t_{k-1} + \frac{1}{2}(t_k - t_{k-1}))$ we would've got the Stratonovich integral.

The advantage in the choice of t_i to be the left extreme of the interval is that we don't require to know the future of the process, only the present time.

Formally, the definition of the Itô integral is done via random step functions f with $\int_a^b \mathbb{E}(f)^2 < \infty$ (when this happens we say that $f \in \mathcal{L}^2$), and using a density argument to extend it to random functions in $\mathcal{L}^2$. Then the Itô integral

$$I(f) = \int_a^b f(t)\, dW_t,$$

can be defined as a linear mapping $I : \mathcal{L}^2 \to L^2$. The next theorem summarises this construction.

Theorem 4.3. *The integral I defined for random step functions f in $\mathcal{L}^2$ as*

$$I(f) = \sum_{k=1}^{n} f(t_{k-1})(W_{t_k} - W_{t_{k-1}}),$$

extends to a continuous linear random functional from $\mathcal{L}^2$ into L^2 which satisfies

$$\mathbb{E}(I(f)) = 0,$$
$$\|I(f)\|_2 = \|f\|.$$

Note that the properties given in this last theorem are the extension of those in Theorem 4.2 to the entire space $\mathcal{L}^2$ (the same can be done for more general spaces, but this will be sufficient for us).

4.2. **Properties of the stochastic integral.** Throughout this section we will assume that f is some random function in $\mathcal{L}^2$.

For any disjoint Borel sets W_1 and W_2,

$$\int_{W_1 \cup W_2} f(t)\, dW_t = \int_{W_1} f(t)\, dW_t + \int_{W_2} f(t)\, dW_t.$$

In particular, if $a \leqslant t_1 \leqslant t_2 \leqslant b$,

$$\int_a^{t_2} f(t)\, dW_t = \int_a^{t_1} f(t)\, dW_t + \int_{t_1}^{t_2} f(t)\, dW_t,$$

and setting $X_t = \int_a^t f(s)\, dW_s$ it's clear that X_t is a Markov process. It is also a martingale since by independence of the increments with respect to $\mathscr{F}_t$, $\mathbb{E}(X_{t+s} - X_t | \mathscr{F}_t) = \mathbb{E}(X_{t+s} - X_t) = 0$.

Proposition 4.4. *Let $f \in \mathcal{L}^2$ and $X_t = \int_a^t f(s)\, dW_s$. Then for any $r > 0$,*

$$P\Big(\sup_{[a,b]} |X_t| > r\Big) \leqslant r^{-2}\mathbb{E}\bigg[\int_a^b f(t)^2\, dt\bigg],$$

and

$$\mathbb{E}\bigg[\sup_{[a,b]} |X_t|^2\bigg] \leqslant 4\mathbb{E}\bigg[\int_a^b f(t)^2\, dt\bigg].$$

4.3. **Itô isometry and a generalisation.** Itô isometry is a crucial fact in Itô calculus. Let W_t be the standard Brownian motion and X_t be a stochastic process adapted to the natural filtration of Brownian motion. Then

$$\mathbb{E}\bigg[\Big(\int_0^t X_s\, dW_s\Big)^2\bigg] = \mathbb{E}\bigg[\int_0^t X_s^2\, ds\bigg].$$

This formula is generalised in the following theorem by Burkholder–Davis–Gundy. The constant $(10)^p$ follows from [Peš96, (2.5) and (2.23)].

Theorem 4.5 ([BDG72]). *Let X_t be a real-valued W_t integrable adapted process. Then for all $1 \leqslant p < \infty$,*

$$\mathbb{E}\Big[\Big(\sup_{0\leqslant t\leqslant 1}\Big|\int_0^t X_s\,dW_s\Big|\Big)^{2p}\Big] \leqslant (10p)^p\,\mathbb{E}\Big[\Big(\int_0^1 X_s^2\,ds\Big)^p\Big].$$

4.4. **Itô's formula.** Let $f, g \in \mathcal{L}^2$ (this can be done more generally, e.g. we don't need $\int(\mathbb{E}f)^2 < \infty$) with $f \in L^1[a,b]$. Then the equation

$$X_t = X_a + \int_a^t f(s)\,ds + \int_a^t g(s)\,dW_s$$

defines a stochastic process with continuous sample paths a.s.. Written in differential form

$$dX_t = f(t)\,dt + g(t)\,dW_t. \tag{4.5}$$

By Theorem 4.3, since $X_t - X_s = \int_s^t f(u)\,du + \int_s^t g(u)\,dW_u$,

$$\mathbb{E}(X_t - X_s) = \int_s^t f(u)\,du,$$

and

$$\begin{aligned}\mathbb{E}\big[(X_t - X_s)^2\big] &= \mathbb{E}\Big[\Big(\int_s^t f(u)\,du\Big)^2\Big] + \mathbb{E}\Big[\Big(\int_s^t g(u)\,dW_u\Big)^2\Big] \\ &\quad - 2\mathbb{E}\Big[\Big(\int_s^t f(u)\,du\Big)\Big(\int_s^t g(u)\,dW_u\Big)\Big] \\ &= \Big[\int_s^t f(u)\,du\Big]^2 + \mathbb{E}\Big[\Big(\int_s^t g(u)\,dW_u\Big)^2\Big] - 2\Big(\int_s^t f(u)\,du\Big)\mathbb{E}\Big[\int_s^t g(u)\,dW_u\Big],\end{aligned}$$

by Theorem 4.3, and using the fact that $\mathbb{E}\Big[\big(\int_a^b Y_u\,dW_s\big)^2\Big] = \mathbb{E}\big[\int_a^b Y_u^2\,du\big]$, we get

$$\mathbb{V}(X_t - X_s) = \int_s^t g(u)^2\,du$$

If $F(t,x)$ is a sufficiently smooth deterministic function defined for all $t \in [a,b]$ and X_t is a process with stochastic differential (4.5). Then $F(t, X_t)$ determines a process with stochastic differential

$$dF(t, X_t) = \widetilde{f}(t, X_t)\,dt + \widetilde{g}(t, X_t)\,dW_t. \tag{4.6}$$

Itô's formula gives analytic expressions for $\widetilde{f}$ and $\widetilde{g}$ in terms of the partial derivatives of F and the functions f and g.

Theorem 4.6. *Let $X = \{X_t : 0 \leqslant t < \infty\}$ be a continuous martingale. Let $F : [0,T] \times \mathbb{R} \to \mathbb{R}$ be a continuous function with continuous partial derivatives $\partial F/\partial t$, $\partial F/\partial x$ and $\partial^2 F/\partial x^2$. Then the process $F(t, X_t)$ has a stochastic differential*

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

In particular, if X_t has stochastic differential (4.5), then

$$dF(t, X_t) = \frac{\partial F}{\partial t}(t, X_t)\,dt + f(t)\frac{\partial F}{\partial x}(t, X_t)\,dt + \frac{1}{2}g(t)^2\frac{\partial^2 F}{\partial x^2}(t, X_t)\,dt + g(t)\frac{\partial F}{\partial x}(t, X_t)\,dW_t,$$

and

$$F(t,X_t)-F(0,X_0)=\int_0^t \frac{\partial F}{\partial t}(x,X_s)+f(s)\frac{\partial F}{\partial x}(s,X_s)+\frac{1}{2}g(s)^2\frac{\partial^2 F}{\partial x^2}(s,X_s)\,ds+\int_0^t g(s)\frac{\partial F}{\partial x}(s,X_s)\,dW_s.$$

Itô's formula gives

$$dF=\frac{\partial F}{\partial t}\,dt+\frac{\partial F}{\partial x}\,dX+\frac{1}{2}g^2\frac{\partial^2 F}{\partial x^2}\,dt,$$

where the first two terms are what we expect from classical calculus, the third term is the new addition. The reason for this is that if we have $dx = f\,dt + g\,dB$, it is not independent of dt, and so

$$(dx)^2 = f^2(dt)^2 + 2fg\,dt\,dB + g^2(dB)^2.$$

The key point is that $(dB)^2$ behaves like dt in mean square calculus.

4.5. **A note on quadratic variation.** Here I've tried to maintain a simple argument, but these things can be done in far more generality. In what I wrote before we depend a lot on 'the fact' that $(dW_t)^2 = dt$, but what does that mean? And how can we define all these things when we are not working with something so nice as classical Brownian motion?

We know that one of the problems is that the paths of Brownian motion are of unbounded variation, and its because of that, that we choose to work with the quadratic variation (see (4.3)). I'm writing this section only to be aware of the notation that I'll introduce and to also know that all these things can be done in more general settings. What is more, this is the way that stochastic integration *should* be defined, and is used, for example, in [KS91].

Let X_t be a stochastic process. Fix $t > 0$ and let $\Pi = \{t_0, t_1, \dots, t_n\}$ be a partition of $[0,t]$. Define the p-th variation of X over the partition Π to be

$$\mathcal{V}_t^{(p)}(\Pi)=\sum_{k=1}^{m}|X_{t_k}-X_{t_{k-1}}|^p.$$

Then, in probability,

$$\lim_{\|\Pi\|\to 0}\mathcal{V}_t^{(2)}(\Pi)=\langle X\rangle_t.$$

Sometimes the quadratic variation $\langle X\rangle_t$ is defined as the unique, adapted, natural, increasing, process, for which $\langle X\rangle_0 = 0$ and $X^2-\langle X\rangle$ is a martingale, but these two definitions are equivalent.

So that's why sometimes Itô's formula is stated as follows (see [KS91, Theorem 3.3.3]): If X_t is a continuous semimartingale with decomposition $X_t = X_0 + M_t + W_t$, then

$$F(X_t)=F(X_0)+\int_0^t F'(X_s)\,dM_s+\int_0^t F'(X_s)\,dW_s+\frac{1}{2}\int_0^t F''(X_s)\,d\langle M\rangle_s.$$

Thus, if $M_t = W_t$, $\langle M\rangle_s = ds$.

If we consider fractional Brownian motion with Hurst parameter $H \in (0,1)$, W_t^H, (recall that classical Brownian motion has $H = 1/2$), then fixing $t > 0$ and letting $t_k^n = kt/n$,

$$\lim_{n\to\infty}\sum_{k=0}^{n-1}|W^H_{t^n_{k+1}}-W^H_{t^n_k}|^p=\begin{cases}+\infty\,, & p<1/H;\\ t\mathbb{E}\big[|W_1^H|^{1/H}\big]\,, & p=1/H;\\ 0\,, & p>1/H.\end{cases}$$

Setting $H = 1/2$ we get the previous result for classical Brownian motion.

4.6. **Examples.**

Example 4.7. *Going back to Section 4.1 we can take $X_t = W_t$ by setting $f = 0$ and $g = 1$ in the definition of the Itô process, and $F(t, x) = x^2/2$ (because it's convenient). Applying Itô's formula we get*

$$d\Big(\frac{W_t^2}{2}\Big) = \frac{\partial F}{\partial x}(t, W_t)\, dW_t + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t, W_t)\, dt.$$

Here we can see the convenience of taking $F(t, x) = x^2/2$, the first term in the right-hand side is the one we want to calculate. Isolating it and changing to integral notation gives

$$\int W_t\, dW_t = \int d\Big(\frac{W_t^2}{2}\Big) - \frac{1}{2}\int dt = \frac{W_t^2}{2} - \frac{t}{2}.$$

As we wanted to show.

Example 4.8. *Let's now calculate $\int_a^b W_t^n\, dW_t$. For this, we will use Theorem 4.6 with $X_t = W_t$, $f = 0$, $g = 1$ and $F(t, x) = x^{n+1}$. And*

$$W_b^{n+1} - W_a^{n+1} = \frac{1}{2}(n+1)n\int_a^b W_t^{n-1}\, dt + (n+1)\int_a^b W_t^n\, dW_t.$$

Isolating the third term gives

$$\int_a^b W_t^n\, dW_t = \frac{1}{n+1}(W_b^{n+1} - W_a^{n+1}) - \frac{n}{2}\int_a^b W_t^{n-1}\, dt$$

Example 4.9 (Integration by parts)**.** *We will now prove that*

$$tW_t = \int_0^t W_s\, dW_s + \int_0^t s\, dW_s.$$

Let $X_t = W_t$ and $F(t, x) = tx$, then by Itô's formula (Theorem 4.6),

$$tW_t = \int_0^t W_s\, ds + \int_0^t s\, dW_s,$$

as we wanted to show.

Example 4.10. *Let $F(t, x) = e^{ax - \frac{a^2}{2}t}$, we will apply Itô's formula to show that $F(t, W_t)$ solves the system*

$$dX_t = 1 + aX_t\, dW_t.$$

Set $X_t = W_t$ in Theorem 4.6, then $f = 0$ and $g = 1$, and

$$\begin{aligned}
F(t, W_t) &= F(0, W_0) + \int_0^t \frac{\partial F}{\partial t}(x, X_s) + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(s, X_s)\, ds + \int_0^t \frac{\partial F}{\partial x}(s, X_s)\, dW_s. \\
&= 1 - \frac{a^2}{2}\int_0^t e^{aW_s - \frac{a^2}{2}s}\, ds + \frac{a^2}{2}\int_0^t e^{aW_s - \frac{a^2}{2}s}\, ds + a\int_0^t e^{aW_s - \frac{a^2}{2}s}\, dW_s \\
&= 1 + a\int_0^t e^{aW_s - \frac{a^2}{2}s}\, dW_s \\
&= 1 + a\int_0^t F(s, W_s)\, dW_s,
\end{aligned}$$

as we wanted to show.

Example 4.11. *Let W_t be standard Brownian motion. Define the process Y by $Y_t = t^2 W_t^3$ for $t \geqslant 0$. Then Y satisfies the stochastic differential equation*

$$dY_t = \Big(\frac{2Y_t}{t} + 3(t^4 Y_t)^{1/3}\Big)\,dt + 3(tY_t)^{2/3}\,dW_t, \quad Y_0 = 0.$$

That Y satisfies $Y_0 = 0$ is trivial. Now to verify that Y satisfies that equation note that $Y = F(t, W_t)$ with $F(t, x) = t^2 x^3$, which is continuous and with continuous second derivatives. Thus, we can use Itô's formula to get (with $f = 0$ and $g = 1$)

$$\begin{aligned} dY_t = dF(t, W_t) &= \frac{\partial F}{\partial t}(t, W_t)\,dt + \frac{1}{2}\frac{\partial^2 F}{\partial x^2}(t, W_t)\,dt + \frac{\partial F}{\partial x}(t, W_t)\,dW_t \\ &= 2tW_t^3\,dt + 3t^2 W_t\,dt + 3t^2 W_t^2\,dW_t \\ &= \Big(\frac{2Y_t}{t} + 3(t^4 Y_t)^1/3\Big)\,dt + 3(tY_t)^{2/3}. \end{aligned}$$

4.6.1. *Hermite polynomials.* We will see that Hermite polynomials play the same role that $t^n/n!$ plays in classic calculus.

For $n \in \mathbb{N}_0$ the n-th *Hermite polynomial* is defined as

$$h_n(x, t) := \frac{(-t)^n}{n!} e^{\frac{x^2}{2t}} \frac{d^n}{dx^n}\Big(e^{-\frac{x^2}{2t}}\Big).$$

Then

$$\begin{cases} h_0(x,t) = 1, \ h_1(x,t) = x, \\ h_2(x,t) = \frac{x^2}{2} - \frac{t}{2}, \ h_3(x,t) = \frac{x^3}{6} - \frac{tx}{2}, \\ h_4(x,t) = \frac{x^4}{24} - \frac{tx^2}{4} + \frac{t^2}{8}, \ \text{etc...} \end{cases}$$

For $t \geqslant 0$ and $n \in \mathbb{N}_0$,

$$\int_0^t h_n(W_s, s)\,dW_s = h_{n+1}(W_t, t).$$

Equivalently,

$$dh_{n+1}(W_t, t) = h_n(W_t, t)\,dW_t.$$

First note that

$$\frac{d^n}{d\lambda^n}\Big(e^{\lambda x - \frac{\lambda^2 t}{2}}\Big)|_{\lambda=0} = (-t)^n e^{\frac{x^2}{2t}} \frac{d^n}{dx^n}\big(e^{-\frac{x^2}{2t}}\big) = n! h_n(x, t).$$

Thus, $e^{\lambda x - \frac{\lambda^2 t}{2}} = \sum_{n=0}^\infty \lambda^n h_n(x, t)$. Define the stochastic process Y_t as

$$Y_t = e^{\lambda W_t - \frac{\lambda^2 t}{2}} = \sum_{n=0}^\infty \lambda^n h_n(W_t, t).$$

From Example 4.10 we know that Y_t solves the equation

$$Y_t = 1 + \lambda \int_0^t Y_s\,dW_s.$$

That is,

$$\begin{aligned} \sum_{n=0}^\infty \lambda^n h_n(W_t, t) &= 1 + \lambda \int_0^t \sum_{n=0}^\infty \lambda^n h_n(W_s, s)\,dW_s \\ &= 1 + \sum_{n=1} \lambda^n \int_0^t h_{n-1}(W_s, s)\,dW_s. \end{aligned}$$

Since this holds for any λ, then

$$h_n(W_t, t) = \int_0^t h_{n-1}(W_s, s)\, dW_s.$$

5. d-dimensional Brownian motion and higher dimensional Itô calculus

The standard d-dimensional Brownian motion $\boldsymbol{W}_t = (W_t^1, \ldots, W_t^d)$, defined for $t \geqslant 0$, has $\mathbb{R}^d$ as its state space, where for each $\ell = 1, \ldots, d$, W_t^ℓ is an independent scalar standard Brownian motion as defined previously. That is, each W_t^ℓ for $\ell = 1, \ldots, d$ is a scalar process with independent, stationary, $N(0, |t-s|)$-distributed increments $W_t^\ell - W_s^\ell$, and satisfying that $W_t 0^\ell = 0$ almost surely.

Definition 5.1. *We say that an $\mathbb{R}^n$ valued stochastic process $\boldsymbol{X}_t$ is in $\mathcal{L}_n^2(0,T)$ if $\boldsymbol{X}_t = (X_t^i)_{i=1}^n$ and for each $i = 1, \ldots, n$,*

$$\mathbb{E}\bigg[\int_0^T (X_t^i)^2\, dt\bigg] < \infty.$$

That is, if each $X_t^i \in \mathcal{L}^2$.

Given an n-dimensional random variable $\boldsymbol{X}_0$ independent of $\boldsymbol{W}_t$, we will take $\mathscr{F}_t \coloneqq \mathscr{F}(\boldsymbol{W}_s(0 \leqslant s \leqslant t), \boldsymbol{X}_0)$, the σ-algebra generated by the history of $\boldsymbol{W}_t$ and $\boldsymbol{X}_0$.

Let $\boldsymbol{f} : \mathbb{R}^n \times [0,T] \to \mathbb{R}^n$ and $\boldsymbol{g} : \mathbb{R}^n \times [0,T] \to \mathbb{R}^d$ be two deterministic functions. We say that $\boldsymbol{X}_t$ is a solution of the Itô stochastic differential equation

$$d\boldsymbol{X}_t = \boldsymbol{f}\, dt + \boldsymbol{g}\, d\boldsymbol{W}_t, \tag{5.1}$$

if

$$\boldsymbol{X}_t = \boldsymbol{X}_s + \int_s^t \boldsymbol{f}(u, \boldsymbol{X}_u)\, du + \int_s^t \boldsymbol{g}(u, \boldsymbol{X}_u)\, d\boldsymbol{W}_u,$$

which means that for each $i = 1, \ldots, n$

$$dX_t^i = f^i\, dt + \sum_{\ell=1}^d g^{i\ell}\, dW_t^\ell.$$

Theorem 5.2 (Itô's chain rule in $\boldsymbol{n}$ dimensions)**.** *Let $\boldsymbol{X}_t$ be an n-dimensional continuous martingale. Let $F : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ be a continuous function with continuous partial derivatives $\frac{\partial F}{\partial t}, \frac{\partial F}{\partial x_i}, \frac{\partial^2 F}{\partial x_i \partial x_j}$ for $i, j = 1, \ldots, n$. Then the process*

$$dF(t, \boldsymbol{X}_t) = \frac{\partial F}{\partial t}(t, \boldsymbol{X}_t)\, dt + \sum_{i=1}^n \frac{\partial F}{\partial x_i}(t, \boldsymbol{X}_t) dX_t^i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 F}{\partial x_i \partial x_j}(t, \boldsymbol{X}_t) \sum_{\ell=1}^d g^{i\ell} g^{j\ell}\, dt.$$

In particular, if $\boldsymbol{X}_t$ has stochastic differential (5.1), then

$$\begin{aligned} dF(t, \boldsymbol{X}_t) = {} & \frac{\partial F}{\partial t}(t, \boldsymbol{X}_t)\, dt + \sum_{i=1}^n f^i(t, \boldsymbol{X}_t) \frac{\partial F}{\partial x_i}(t, \boldsymbol{X}_t)\, dt \\ & + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 F}{\partial x_i \partial x_j}(t, \boldsymbol{X}_t) \sum_{\ell=1}^d g^{i\ell}(t, \boldsymbol{X}_t) g^{j\ell}(t, \boldsymbol{X}_t)\, dt \\ & + \sum_{i=1}^n \sum_{\ell=1}^d g^{i\ell}(t, \boldsymbol{X}_t) \frac{\partial F}{\partial x_i}(t, \boldsymbol{X}_t)\, dW_t^\ell, \end{aligned}$$

and

$$\begin{aligned}F(t,\boldsymbol{X}_t)-F(0,\boldsymbol{X}_0)=\int_0^t \frac{\partial F}{\partial t}(s,\boldsymbol{X}_s)+\sum_{i=1}^n f^i(s,\boldsymbol{X}_s)\frac{\partial F}{\partial x_i}(s,\boldsymbol{X}_s)\\ +\frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 F}{\partial x_i\partial x_j}(s,\boldsymbol{X}_s)\sum_{\ell=1}^d g^{i\ell}(s,\boldsymbol{X}_s)g^{j\ell}(s,\boldsymbol{X}_s)\,ds\\ +\sum_{i=1}^n\sum_{\ell=1}^d\int_0^t g^{i\ell}(s,\boldsymbol{X}_s)\frac{\partial F}{\partial x_i}(s,\boldsymbol{X}_s)\,dW_s^\ell.\end{aligned}$$

References

[BDG72] D. L. Burkholder, B. J. Davis, and R. F. Gundy. Integral inequalities for convex functions of operators on martingales. *Berkely Symp. on Math. Statist. and Prob.*, 6.2:223–240, 1972.

[Eva12] L.C. Evans. *An Introduction to Stochastic Differential Equations.* American Mathematical Society, 2012.

[Gar88] T.C. Gard. *Introduction to Stochastic Differential Equations.* Monographs and Textbooks in Pure and Applied Mathematics. New York, NY, 1988.

[KS91] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus.*, volume 113 of *Grad. Texts Math.* New York etc.: Springer-Verlag, 2nd ed. edition, 1991.

[Peš96] G. Peškir. On the exponential Orlicz norms of stopped Brownian motion. *Studia Mathematica*, 117(3):253–273, 1996.

A. E. de Orellana, University of St Andrews, Scotland

Email address: aedo1@st-andrews.ac.uk