

# Stochastic calculus

Gabriele Pinna

October 16, 2024

## Contents

<b>1</b>	<b>Introduction: Conditional expectation and Martingales</b>	<b>3</b>
1.1	Conditional expectation . . . . .	3
1.2	Martingales . . . . .	4
1.2.1	Examples . . . . .	4
1.3	Markov process . . . . .	4
1.4	Discrete stochastic integral . . . . .	4
1.4.1	Martingale betting strategy . . . . .	5
1.4.2	Optional stopping time theorem . . . . .	5
1.4.3	Martingale convergence theorem . . . . .	5
<b>2</b>	<b>Wiener process</b>	<b>5</b>
2.1	Definition and properties . . . . .	6
2.1.1	List of properties . . . . .	6
2.2	Arcsine and Levy distribution . . . . .	7
2.3	Existence . . . . .	8
2.4	Quadratic variation and covariation . . . . .	8
2.4.1	Higher dimensions . . . . .	8
<b>3</b>	<b>Stochastic calculus</b>	<b>9</b>
3.1	Introduction . . . . .	9
3.2	Itô integration . . . . .	9
3.2.1	Simple processes . . . . .	9
3.2.2	Itô formula . . . . .	12
3.2.3	Stochastic product rule . . . . .	12
3.2.4	Examples . . . . .	12
3.3	Kolmogorov equations . . . . .	13
3.3.1	Markov semigroup . . . . .	13
3.3.2	Diffusion processes . . . . .	15
3.3.3	Kramers-Moyal expansion . . . . .	16
3.4	Stratonovich integration . . . . .	17
<b>4</b>	<b>Important (continous) processes</b>	<b>18</b>
4.1	Brownian bridge . . . . .	18
4.2	Geometric Brownian motion . . . . .	19
4.3	Ornstein-Uhlenbeck process . . . . .	20
4.4	Cox-Ingersoll-Ross model . . . . .	21

---

<b>5</b>	<b>Path integral</b>	<b>21</b>
5.1	Feynman-Kac formula . . . . .	21
5.2	Black-Scholes equation . . . . .	22
<b>6</b>	<b>Jump processes</b>	<b>23</b>
<b>7</b>	<b>Fractional Brownian motion</b>	<b>23</b>

# 1 Introduction: Conditional expectation and Martingales

This is a summary on stochastic calculus following mostly:

- “*Introduction to stochastic calculus with applications*”
- “*Stochastic processes and applications*” by G.A. Pavliotis.

First we introduce some useful definitions such as conditional expectations and martingales which are stochastic processes that model fair games.

## 1.1 Conditional expectation

Usually, we think of conditional expectation as a non-random variable. Given to random variables  $X$  and  $Y$  let  $f_{X,Y}(x,y)$  be the joint distribution, we define the conditional distribution as  $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$  from which we have Bayes’ theorem

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} . \quad (1)$$

We define the conditional expectation as

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}} dx x f_{X|Y}(x|y) \quad (2)$$

this is a number which can be thought of as the best guess for the random variable  $X$  given the information that another random variable  $Y$  takes the value  $y$ . We can turn this expectation into a random variable, which reflects the outcome of an experiment where the random variable  $X$  is drawn subject to a random constraint represented by a random variable  $Y$ . This defines a new random variable which is also called “expectation value”

$$E[X|Y] = \int_{\mathbb{R}} dx x f_{X|Y}(x|Y) \quad (3)$$

this is a random variable that depends on the random variable  $Y$ . We can then calculate the expectation value of this random variable

$$\mathbb{E}_Y[E[X|Y]] = \int_{\mathbb{R}} dy \mathbb{E}[X|Y = y] f(y) = \int_{\mathbb{R}} dy \int_{\mathbb{R}} dx x f_{X|Y}(x|y) f(y) = \mathbb{E}_{X,Y}[X] \quad (4)$$

which is the same as the expectation value of  $X$ ; this is because we are averaging over all possible values of the constrain  $Y$ .

In general, let  $\mathcal{F}_n = X_1, X_2, \dots, X_n$  be a sequence of events (random variables) where the index labels time and  $\mathcal{F}_0$  represent the lack of information. We require the conditional expectation to satisfy the property

- $E[Y|\mathcal{F}_0] = Y$
- $E[Y|\mathcal{F}_n] = f(X_1, \dots, X_n)$

if the second condition is satisfied we say that  $Y$  is  $\mathcal{F}_n$ -measurable which means that the conditional expectation depends only on the information available at time  $n$ . Additionally, it satisfies properties like

- Linearity:  $E[aY_1 + bY_2|\mathcal{F}] = aE[Y_1|\mathcal{F}] + bE[Y_2|\mathcal{F}]$
- Projection: if  $m < n$   $E[E[X|\mathcal{F}_n]|\mathcal{F}_m] = E[X|\mathcal{F}_m]$

## 1.2 Martingales

A martingale is a stochastic process  $\{M_1, M_2, \dots\}$  that satisfies

- For each  $n$   $M_n$  is a  $\mathcal{F}_n$ -measurable with  $\mathbb{E}[|M_n|] < \infty$
- If  $m < n$  we have  $E[M_n|\mathcal{F}_m] = M_m \implies E[M_n - M_m|\mathcal{F}_m] = 0$

the second condition means that the conditional expectation based on information up to time  $m$  is the random variable at time  $m$ . Martingales model fair games; in fact, the expectation value does not depend on what happened in the past. Using the tower property we find for a martingale

$$\mathbb{E}[M_n] = \mathbb{E}[E[M_n|\mathcal{F}_0]] = \mathbb{E}[M_0] \quad (5)$$

Similarly we define, with  $m < n$ ,

- Supermartingale:  $E[M_n|\mathcal{F}_m] \leq M_m$
- Submartingale:  $E[M_n|\mathcal{F}_m] \geq M_m$

a supermartingale models a game that is not convenient, whereas a submartingale models a game that is convenient.

### 1.2.1 Examples

Let  $\{X_j\}_{j=1}^n$  be  $n$  independent random variables with  $\mathbb{E}[X_j] = 0$ , and  $S_n = \sum_{j=1}^n X_j$ , then  $S_n$  is a martingale:  $E[S_n|\mathcal{F}_m] = S_m \forall n > m$ .

Let  $\mathbb{E}[X_j^2] = \sigma_j^2$  and  $A_n = \sum_{j=1}^n \sigma_j^2$ ; then,  $M_n = S_n^2 - A_n$  is a martingale.

## 1.3 Markov process

Another important stochastic process is a Markov process which satisfies the Markov property. Let  $m < n$  and  $\mathcal{F}_m$  be a filtration, then the (discrete) stochastic process  $\{X_j\}$  satisfied the Markov property if

$$P(X_{n+1} = x_{n+1} | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) \quad (6)$$

this implies that the knowledge of the past does not affect the probability.

## 1.4 Discrete stochastic integral

We introduce the concept of stochastic integral with an example. Consider  $\{M_j\}$  to be a martingale where  $M_j$  is the asset price at time-step  $j$  (cumulative). Assume that at each time-step we place a series of bets  $\{B_j\}$ , regarding whether the asset price will go up or down. We allow for negative bet which means that we are betting on the price of the asset to go down. The money won at this game at time-step  $n$  is given by

$$W_n = \sum_{j=1}^n (M_j - M_{j-1})B_j \quad (7)$$

we now proved that under some assumptions  $W_n$  is also a martingale. We assume that  $B_j$  is bounded  $B_j < \infty$  and that it is  $\mathcal{F}_{n-1}$ -measurable; this means that we can adjust the bet based on how well we are doing. It follows that  $\mathbb{E}[W_n] < \infty$ ; next we need to prove that  $E[W_{n+1}|\mathcal{F}_n] = W_n$ .

$$\begin{aligned}
E[W_{n+1}|\mathcal{F}_n] &= E[W_n + (M_{n+1} - M_n)B_{n+1}|\mathcal{F}_n] = E[W_n|\mathcal{F}_n] + E[(M_{n+1} - M_n)B_{n+1}|\mathcal{F}_n] \\
&= W_n + B_{n+1}E[(M_{n+1} - M_n)|\mathcal{F}_n] = W_n
\end{aligned}$$

Hence,  $W_n$  is a martingale; thus, we are not beating the fair game in a finite amount of time. The quantity  $W_n$  is a discrete stochastic integral.

#### 1.4.1 Martingale betting strategy

As an aside, consider a coin toss with  $\mathbb{P}[X_j = 1] = \mathbb{P}[X_j = -1] = 1/2$ . We are going to find a strategy to beat this fair game but it requires an infinite amount of time. We consider betting 1 on the first time, if we win we quit otherwise we double the bet on the next game and so on. We can see that if we win at the  $n$ -th turn we make  $2^{n-1}$  but we have lost  $\sum_{j=1}^{n-2} 2^j = (2^{n-1} - 1)/(2 - 1) = 2^{n-1} - 1$ ; thus we can see that we gain 1. With probability one, eventually we will win a game; therefore, we are guaranteed to win 1. In general, we cannot beat a martingale in a finite amount of time but we could in an infinite amount of time.

#### 1.4.2 Optional stopping time theorem

The fact that a martingale cannot be beaten in a finite time is the content of a theorem known as *optional stopping time theorem*. A stopping time  $T$  with respect to a filtration  $\{\mathcal{F}_n\}$  is defined to be a non-negative random variable such that for each  $n$  the event  $T \leq n$  is  $\mathcal{F}_n$ -measurable. In general, the stopping time  $T$  is used when after some time  $T$  has passed we stop betting. For example, let

$$M_{n \wedge T} = M_0 + \sum_{j=1}^n B_j(M_j - M_{j-1}) \quad (8)$$

be a martingale that denotes a betting strategy where  $B_j = 1$  if  $j < T$  or 0 otherwise. The optional sampling theorem says that  $M_{n \wedge T}$  is a martingale ( $\mathbb{E}[M_{n \wedge T}] = \mathbb{E}[M_0]$ ) and if  $T$  is bounded,  $\mathbb{P}[T < k] = 1$ , then

$$\mathbb{E}[M_T] = \mathbb{E}[M_0] \quad (9)$$

#### 1.4.3 Martingale convergence theorem

Let  $M_n$  be a martingale with respect to a filtration  $\mathcal{F}_n$  and  $\mathbb{E}[M_n] \leq C$  then there exists a random variable  $M_\infty$  such that with probability one

$$\lim_{n \rightarrow \infty} M_n = M_\infty \quad (10)$$

## 2 Wiener process

A Wiener process, also called Brownian motion, can be viewed as the limit of random walk as the time and space increments tend to zero. Let's consider an example where  $P[X_j = 1] = P[X_j = -1] = 1/2$  and let  $S_n = \sum_{1 \leq j \leq n} X_j$  be the random walk. We view the label  $j$  as time-steps and the value of  $X_j$  as space which in this case increment by  $\pm 1$ . We can generalise it considering a generic time  $t = N\Delta t$  and a generic space-increment  $\Delta x$ , after time  $t$  the total space increment is

$$W_n = \Delta x S_n \quad (11)$$

we would like to scale  $\Delta x$  such that  $\text{Var}[W_n] = 1$ , i.e.  $(\Delta x)^2 = n/2 = N\Delta t/2$ . Hence, we need  $\Delta x \sim \sqrt{\Delta t}$ . We now present a rigorous definition.

## 2.1 Definition and properties

A Wiener process is a stochastic process  $B_t$  (sometimes denoted  $W_t$ ) such that

- **Stationary increments.** If  $s < t$ , then the distribution of  $B_t - B_s$  is the same as that of  $B_{t-s} - B_0$ .
- **Independent increments.** If  $s < t$ , the random variable  $B_t - B_s$  is independent of the values  $B_r$  for  $r \leq s$ .
- **Continuous paths.** The function  $t \mapsto B_t$  is a continuous function of  $t$ .

A stochastic process with stationary and independent increments is known as Levy process. One can prove, that if  $B_t$  is a process satisfying the three conditions above, then the distribution of  $B_t$  for each  $t$  must be normal  $\mathcal{N}(\mu, \sigma^2)$ . This is done

Using the properties we obtain

$$\mathbb{E}[B_t] = \mathbb{E}[B_s] + \mathbb{E}[B_t - B_s] = \mathbb{E}[B_s] + \mathbb{E}[B_{t-s}] \implies \mathbb{E}[B_t] = t\mu \quad (12)$$

and similarly  $\text{Var}[B_t] = t\sigma^2$ . An important property is the self-similarity: if  $W_t$  is a Wiener process so is  $1/\sqrt{c}W_{ct}$ . It follows from the fact that if  $X \sim \mathcal{N}(0, 1)$  then  $\sigma X + M \sim \mathcal{N}(M, \sigma^2)$ . It is customary to define the standard Brownian motion with mean 0 and variance 1 and  $B_0 = 0$  almost surely.

Using these definitions we can calculate  $\mathbb{E}[W_t W_s]$  assuming  $s > t$  we get  $\mathbb{E}[W_t W_s] = \mathbb{E}[W_t(W_t + W_s - W_t)] = \mathbb{E}[(W_t)^2] \mathbb{E}[W_t(W_s - W_t)]$  since  $s > t$   $W_t$  is independent of  $W_s - W_t$  and thus  $\mathbb{E}[W_t(W_s - W_t)] = \mathbb{E}[W_t] \mathbb{E}[(W_s - W_t)] = 0$  and we thus get  $\mathbb{E}[W_t W_s] = \min(t, s)$ .

We can show that  $(W_t)^2 - t$  is a martingale. Assume  $s < t$

$$E[(W_t)^2 | \mathcal{F}_s] = E[(W_s)^2 | \mathcal{F}_s] + E[(W_t - W_s)^2 | \mathcal{F}_s] + 2E[W_s(W_t - W_s) | \mathcal{F}_s] = (W_s)^2 + t - s \quad (13)$$

thus

$$E[(W_t)^2 - t | \mathcal{F}_s] = (W_s)^2 - s. \quad (14)$$

We used the fact that  $W_t - W_s$  is independent of the information  $\leq s$  thus  $E[(W_t - W_s)^2 | \mathcal{F}_s] = E[(W_t - W_s)^2] = E[(W_{t-s})^2] = t - s$  where the second equality follows from stationarity. Note that we will see that there is an alternative way to prove this using Ito's formula: a stochastic process  $f(t, W_t)$  is a martingale iff  $df/dt = -1/2 df^2/dW_t$ . We can see that for  $f(t, W_t) = W_t^2 - t$  we have that both sides are equal to  $-1$ .

We can also check that  $W_t$  is no-where differentiable, since  $W_{t+dt} - W_t = \sqrt{dt}\eta$  where  $\eta$  is a Gaussian variables distributed as  $\mathcal{N}(0, 1)$  (since the difference must be distributed as  $\mathcal{N}(0, dt)$ ). We then have that  $\lim_{dt \rightarrow 0} (W_{t+dt} - W_t)/dt = \infty$ .

$W_t$  is also a Markov process because, for  $s > 0$ ,  $W_{t+s} - W_t$  is independent of  $\{W_k\}_{k \leq t}$ , thus  $E[W_{t+s} - W_t | \mathcal{F}_t] = E[W_{t+s} - W_t | W_t]$ .

### 2.1.1 List of properties

We summarise some properties of Wiener processes

- $\mathbb{E}[W_t W_s] = \min(s, t)$
- $W_t \sim \mathcal{N}(0, t)$
- $W_t$  is a continuous martingale  $E[W_t | \mathcal{F}_s] = W_s \forall s \leq t$
- $(W_t)^2 - t$  is a martingale
- $W_t$  is nowhere differentiable

- $W_t$  is a Markov process
- Reflection principle:  $\mathbb{P}[\max_{0 \leq s \leq t}(B_s) \geq a] = 2\mathbb{P}[B_t > a] = 2 \int_{-\infty}^a dx \frac{\exp(-\frac{x^2}{2t})}{\sqrt{2\pi t}}$

## 2.2 Arcsine and Levy distribution

An important concept is that of stopping time.

First consider the hitting time  $\tau = \inf\{t | W_t = x\}$ , by **reflection principle** we have

$$\mathbb{P}[\tau \leq t] = 2\mathbb{P}[W_t \geq x] \quad (15)$$

This is proved by noticing that  $\mathbb{P}[\tau \leq t, W_t > x] = \mathbb{P}[\tau \leq t, W_t < x]$  since after the hitting time  $W_t$  can either be more than  $x$  or less than  $x$  with equal probability. Thus

$$\mathbb{P}[\tau \leq t] = \mathbb{P}[\tau \leq t, W_t > x] + \mathbb{P}[\tau \leq t, W_t < x] \quad (16)$$

$$= 2\mathbb{P}[\tau \leq t, W_t > x] = 2\mathbb{P}[\tau \leq t, W_t \geq x] = 2\mathbb{P}[W_t \geq x] \quad (17)$$

where  $\mathbb{P}[\tau \leq t, W_t \geq x] = \mathbb{P}[W_t \geq x]$  because the condition  $\tau \leq t$  is automatically satisfied if  $W_t \geq x$ . The reflection principle can be rephrased as: the reflected stochastic process

$$\tilde{W}_t = \begin{cases} W_t & t < \tau \\ 2x - W_t & t \geq \tau \end{cases} \quad (18)$$

is a Wiener process.

Consider the following stopping time  $T = \operatorname{argmax}_{t \in [0,1]}(W_t)$ , we will prove that it follows an arcsine distribution. Let  $S_t = \sup_{t \in [0,1]} W_t$

$$\mathbb{P}[T \leq t] = \mathbb{P}[S_t > S_{1-t}] \quad (19)$$

since  $S_t \sim \mathcal{N}(0, t)$  and  $S_{1-t} \sim \mathcal{N}(0, 1-t)$  we can rescale  $S_t = \sqrt{t}X$  and  $S_{1-t} = \sqrt{1-t}Y$  where  $X, Y$  are standard Gaussian variables. The constrain can be re-written as

$$\sqrt{t}X > \sqrt{1-t}Y \implies \frac{|Y|}{\sqrt{X^2 + Y^2}} \equiv \sin(\theta) < \sqrt{t} \quad (20)$$

The variable  $\theta$  is uniformly distributed along  $[-\pi, \pi]$ , thus

$$\mathbb{P}[-\pi < \theta < \alpha] = \frac{\alpha + \pi}{2\pi} \implies \mathbb{P}[|\sin(\theta)| < \sqrt{t}] = \frac{2}{\pi} \arcsin(\sqrt{t}) \quad (21)$$

Furthermore, we can also prove that  $\tau$  is a random variable distributed according to a Levy distribution

$$\tau_x = \operatorname{argmin}_{t \geq 0}(W_t = x) \sim f(v; x) = \frac{x}{\sqrt{2\pi v^3}} \exp(-\frac{x^2}{2v}) \mathbb{1}_{v \geq 0} . \quad (22)$$

This is proved using the reflection principle

$$\mathbb{P}[\tau_x < t] = 2\mathbb{P}[W_t > x] = 2 \int_x^\infty ds \frac{e^{-s^2/2t}}{\sqrt{2\pi t}} = \int_0^t dy f(y) \quad (23)$$

to find  $f(y)$  we can just take one detivative which yields

$$f(t) = 2\partial_t \int_x^\infty ds \frac{e^{-s^2/2t}}{\sqrt{2\pi t}} = 2 \int_x^\infty ds \frac{1}{2} \partial_s^2 \frac{e^{-s^2/2t}}{\sqrt{2\pi t}} = -\partial_x \frac{e^{-x^2/2t}}{\sqrt{2\pi t}} = \frac{x e^{-\frac{x^2}{2t}}}{\sqrt{2\pi t^3}} \quad (24)$$

which implies that the hitting time  $\tau_x$  is distributed according to a Levy distribution.

We can also prove that  $L = \sup\{t \in [0, 1] | W_t = 0\}$  is distributed according to an arcsine distribution.  $L$  represents the maximum time  $t$  when the Wiener process hits zero. The probability that  $L < t$  is equal to the probability that a Wiener process initialised at  $W_t$  does not hit zero in the remaining time  $1 - t$ :  $\mathbb{P}[L < t] = \mathbb{P}[\tau_{W_t} > 1 - t]$ . Hence

$$\int_{-\infty}^{\infty} dy \phi(y, t) \mathbb{P}[\tau_y > 1 - t] = \int_{-\infty}^{\infty} dy \phi(y, t) \int_{1-t}^{\infty} dv \frac{ye^{-\frac{y^2}{2v}}}{\sqrt{2\pi v^3}} = \frac{2}{\pi} \arcsin(\sqrt{t}) \quad (25)$$

### 2.3 Existence

The existence of a stochastic process satisfying the three conditions can be proved using Wiener isometry. It is based on realising that  $\mathbb{E}[W_t W_s] = \min(s, t)$  is similar to  $\langle \mathbb{1}_{[0,t]}, \mathbb{1}_{[0,s]} \rangle = \min(s, t)$  where  $\langle \cdot, \cdot \rangle$  is an inner product in  $L^2(\mu)$ . We can then construct a map (Wiener isometry)  $I_W$  such that  $I_W(\mathbb{1}_{[0,t]}) = W_t$ . We can then consider a basis in  $L^2(\mu)$ ,  $\{\psi_n\}$  and using the fact that for Gaussian variables  $\mathbb{E}[\eta_j \eta_k] = \delta_{ij}$  we can assign  $I_W(\psi_n) = \eta_n$  thus we have

$$W_t = \sum_{n=1}^{\infty} \eta_n \langle \mathbb{1}_{[0,t]}, \psi_n \rangle \quad (26)$$

### 2.4 Quadratic variation and covariation

Let  $X_j$  be a stochastic process, then we define the **quadratic variation** as

$$\langle X \rangle_t = \lim_{n \rightarrow \infty} \sum_{j=1}^{tn} \left( X_{j/n} - X_{(j-1)/n} \right)^2 \quad (27)$$

For a Wiener process  $W_t$  we have

$$\langle W \rangle_t = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{tn} \left( \frac{W_{j/n} - W_{(j-1)/n}}{\sqrt{1/n}} \right)^2 \quad (28)$$

the quantity in brackets is distributed as  $\mathcal{N}(0, 1)$  hence  $\mathbb{E}[\langle W \rangle_t] = t$  and the variance is  $\text{Var}[\langle W \rangle_t] = \lim_{n \rightarrow \infty} 2/n = 0$  (using  $E[\eta^4] = 3$  for Gaussian variables). Therefore, almost surely  $\langle W \rangle_t = t$ .

This follows from the fact that if  $X$  is a non-negative random variable,  $P(X > 0) = 1$ , and  $E[X] = 0$  then  $P(X = 0) = 1$ . In other words, a non-negative random variable with a zero expected value is almost surely equal to 0. In our case, the random variable is  $(\langle X \rangle_t - E[\langle X \rangle_t])^2$  which is non-negative and it has zero expectation value.

Similarly, we define the **quadratic covariation** between two stochastic processes  $X_t$  and  $Y_t$  as

$$[X, Y]_t = \lim_{n \rightarrow \infty} \sum_{j=1}^{tn} \left( X_{j/n} - X_{(j-1)/n} \right) \left( Y_{j/n} - Y_{(j-1)/n} \right) \quad (29)$$

#### 2.4.1 Higher dimensions

We can trivially generalise the Brownian motion in  $\mathbb{R}^d$  introducing a vector  $W_t = (W_t^1, \dots, W_t^d)$  and we impose that  $W_t$  is distributed as  $\mathcal{N}(0, t\Gamma)$  where  $\Gamma_{ij} = E[\eta_i \eta_j] = \text{Cov}(\eta_j, \eta_i)$  is the covariance matrix. In this case, the quadratic covariation is

$$\langle W_t^i, W_t^j \rangle = t \Gamma_{ij} \quad (30)$$



### 3 Stochastic calculus

#### 3.1 Introduction

Usually in ordinary calculus one wants to solve ordinary differential equations of the form

$$dy = f(y, t)dt \quad (31)$$

which is solved by

$$y(t) = y_0 + \int_0^t ds f(y(s), s) \quad (32)$$

we can solve this analytically (sometimes) or numerically via Euler method by discretising time

$$y((n+1)\Delta t) = y_0 + \Delta t f(y(n\Delta t), n\Delta t) \quad (33)$$

Stochastic calculus introduce randomness in the change

$$dX_t = m(t, X_t)dt + \sigma(t, X_t)dW_t \quad (34)$$

This is a stochastic differential equation (SDE) which should be interpreted as time  $t$ ,  $X_t$  is evolving like a Brownian motion with drift  $m(t, X_t)$  and variance  $\sigma^2(t, X_t)$ . We can still have a discretisation of time

$$X((n+1)\Delta t) = X_0 + \Delta t m(X(n\Delta t), n\Delta t) + \sqrt{\Delta t} \sigma(X(n\Delta t), n\Delta t) \eta \quad (35)$$

where  $\eta \sim \mathcal{N}(0, 1)$  is a Gaussian variable. The formal solution to this SDE is

$$X_t = X_0 + \int_0^t ds m(X_s, s) + \int_0^t dW_s \sigma(X_s, s) \quad (36)$$

where we need to define the stochastic integral  $\int_0^t dW_s \sigma(X_s, s)$

Notice that sometimes in the physics literature it is written as a Langevin equation

$$\dot{X}_t = f(t) + \eta_t \quad (37)$$

and  $\eta_t$  is said to be Gaussian white noise with zero mean  $\mathbb{E}[\eta_t] = 0$  and  $\mathbb{E}[\eta_t \eta_s] = \delta(t - s)$ . Notice that this is just the following statement in disguise  $\mathbb{E}[W_t W_s] = \min(s, t)$  since we have

$$\mathbb{E}[W_t W_s] = \int_0^t dt' \int_0^s ds' \mathbb{E}[dW_{t'} dW_{s'}] = \min(s, t) \implies \mathbb{E}[dW_t dW_s] = \delta(t - s) \quad (38)$$

Therefore, we can think of a Wiener process as the integral of Gaussian white noise.

#### 3.2 Itô integration

In this section we define Itô integration which is a prescription to make sense of the expression  $\int dW_t X_t$ .

##### 3.2.1 Simple processes

In ordinary calculus we have a notion of simple function  $f$  that can be written as a linear combination of indicator functions over some sets  $\{A_k\}$

$$f(x) = \sum_k a_k \mathbb{1}_{A_k} \quad (39)$$

then integration is easily defined as

$$\int f d\mu = \sum_k a_k \mu(A_k) \quad (40)$$

Simple processes play the same role as simple functions

$$A_t = \sum_{k=1}^n \xi_k \mathbb{1}_{[t_{k-1}, t_k)} \quad (41)$$

where  $0 < t_1 < t_2 \cdots < t_n = t$  and  $\xi_k$  is a random variable that is  $\mathcal{F}_{t_k}$ -measurable. We then define

$$\int A_s dW_s = \sum_{k=1}^n \xi_k (W_{t_{i+1}} - W_{t_i}) \quad (42)$$

This is a Itô integral since the value of the integrand is evaluated at the endpoint:  $\xi_i(W_{t_{i+1}} - W_{t_i})$ , if we evaluate it in the middle i.e.  $\xi_i \rightarrow 1/2(\xi_i + \xi_{i+1})$  we have a Stratonovich integral.

This integral over a simple process satisfies the following properties

- **Linearity.** If  $a, b$  are constants, then  $aA_t + bC_t$  is also a simple process, and

$$\int_0^t (aA_s + bC_s) dB_s = a \int_0^t A_s dB_s + b \int_0^t C_s dB_s.$$

Moreover, if  $0 < r < t$ ,

$$\int_0^t A_s dB_s = \int_0^r A_s dB_s + \int_r^t A_s dB_s.$$

- **Martingale property.** The process

$$Z_t = \int_0^t A_s dB_s$$

is a martingale with respect to  $\{\mathcal{F}_t\}$ .

- **Variance rule or Itô isometry.**  $Z_t$  is square integrable, and

$$\text{Var}[Z_t] = \mathbb{E}[Z_t^2] = \int_0^t \mathbb{E}[A_s^2] ds.$$

- **Continuity.** With probability one, the function  $t \mapsto Z_t$  is a continuous function.

We prove the martingale property with  $t_{j+1} = s$

$$E[Z_t | \mathcal{F}_s] = E[Z_s | \mathcal{F}_s] + E\left[\sum_{k=1}^j \xi_k (W_{t_{k+1}} - W_{t_k}) | \mathcal{F}_s\right] \quad (43)$$

$$= Z_s + \sum_{k=1}^j E\left[\xi_k E[(W_{t_{k+1}} - W_{t_k}) | \mathcal{F}_{t_k}] | \mathcal{F}_s\right] \quad (44)$$

$$= Z_s + \sum_{k=1}^j E\left[\xi_k \mathbb{E}[(W_{t_{k+1}} - W_{t_k})] | \mathcal{F}_s\right] = Z_s \quad (45)$$

where we have used the fact that  $\xi_k$  is  $\mathcal{F}_{t_k}$ -measurable and  $W_{t_{k+1}} - W_{t_k}$  is independent of  $\mathcal{F}_{t_k}$ . Notice that the fact that we are using the endpoint  $\xi_k$  allows us to pull it out and then

use the independence property. If we use the Stratonovich interpretation then we cannot pull  $(\xi_{k+1} + \xi_k)/2$  out of the expectation; hence it is not a martingale.

We now show the variance rule

$$\mathbb{E}[Z_t^2] = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[\xi_i(W_{t_{i+1}} - W_{t_i})\xi_j(W_{t_{j+1}} - W_{t_j})] \quad (46)$$

we use the following trick, if  $j > i$  then

$$\mathbb{E}[\xi_i(W_{t_{i+1}} - W_{t_i})\xi_j(W_{t_{j+1}} - W_{t_j})] = \mathbb{E}[E[\xi_i(W_{t_{i+1}} - W_{t_i})\xi_j(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_{t_j}]] \quad (47)$$

notice that  $\xi_i, \xi_j, W_{t_{i+1}} - W_{t_i}$  are all  $\mathcal{F}_{t_j}$ -measurable hence they can be pulled out of  $E[\cdot|\mathcal{F}_{t_j}]$ , whereas  $W_{t_{j+1}} - W_{t_j}$  is independent of  $\mathcal{F}_{t_j}$  hence  $E[W_{t_{j+1}} - W_{t_j}|\mathcal{F}_{t_j}] = \mathbb{E}[W_{t_{j+1}} - W_{t_j}] = \mathbb{E}[W_{t_{j+1}-t_j}] = 0$ . Thus it is zero, similarly for  $i > j$ , the only non-zero term is  $i = j$  which yields

$$\mathbb{E}[Z_t^2] = \sum_{i=1}^n \mathbb{E}[(\xi_i)^2]\mathbb{E}[(W_{t_{i+1}} - W_{t_i})^2] = \sum_{i=1}^n \mathbb{E}[(\xi_i)^2](t_{i+1} - t_i) = \int_0^t ds \mathbb{E}[A_s^2] \quad (48)$$

where the last step follows because  $A_s$  is a simple process (combination of indicator functions). It is clear that we can approximate any continuous process with a sequence of simple process via

$$X_t^{(n)} = \sum_{j=0}^n X_{t_j} \mathbb{1}_{[t_j, t_{j+1})} \quad (49)$$

where  $[0, t]$  has been partitioned in intervals  $t_0 < t_1 < \dots < t_n = t$ . When we make the step size of the partition zero it converges in expectation value to  $X_t$  thus

$$\lim_{n \rightarrow \infty} \int_0^t ds \mathbb{E}[(X_s - X_s^{(n)})^2] = 0 \quad (50)$$

Hence, we define the stochastic integral for a generic continuous stochastic process as

$$\int_0^t dW_s X_s = \lim_{n \rightarrow \infty} \int_0^t dW_s X_s^{(n)} \quad (51)$$

Given an Itô integral  $X_t = \int_0^t dW_s \sigma_s$ , the quadratic variation is given by

$$\langle X \rangle_t = \int_0^t ds \sigma_s^2 \quad (52)$$

the special case of Brownian motion  $W_t = \int_0^t dW_s$  yields  $\langle X \rangle_t = \int_0^t ds = t$  as previously found. Sometimes it is written in differential form:

$$d\langle X \rangle_t = \sigma_t^2 dt \quad (53)$$

Notice that  $W_t$  is distributed as a Gaussian  $W_t \sim \mathcal{N}(0, t)$ , since  $Y_t = \int_0^t f(s) dW_s$  is a linear combination of  $W_t$  we have that  $Y_t$  is also a Gaussian process with  $Y_t \sim \mathcal{N}(0, \int_0^t f(s)^2 ds)$ .

### 3.2.2 Itô formula

Itô formula is the analogous of a Taylor series for stochastic function. We give a first formula valid for Wiener processes

$$df(t, W_t) = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial W_t} dW_t + \frac{1}{2} \frac{\partial^2 f}{\partial W_t^2} dt \quad (54)$$

The notation  $\partial f \partial W_t$  might be confusing, if  $f(t, x)$  is a function then it means  $f^{(0,1)}(t, W_t)$ . Let's consider an example  $f(W_t) = (W_t)^2$  then we have

$$d(W_t^2) = 2W_t dW_t + dt \implies \int_0^t W_s dW_s = \frac{1}{2}(W_t^2 - t) \quad (55)$$

we can see that this does not follow the usual rule of calculus  $\int_0^x ds s = x^2/2$ . Now assume that a stochastic process  $X_t$  satisfies an SDE of the form

$$dX_t = m(t, X_t)dt + \sigma(t, X_t)dW_t \quad (56)$$

then  $f(t, X_t)$  satisfies

$$\begin{aligned} df(t, X_t) &= \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial W_t} dX_t + \frac{1}{2} \frac{\partial^2 f}{\partial X_t^2} d\langle X \rangle_t \\ &= \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial W_t} (m(t, X_t)dt + \sigma(t, X_t)dW_t) + \frac{1}{2} \frac{\partial^2 f}{\partial X_t^2} \sigma^2(t, X_t)dt \\ &= \left[ \frac{\partial f}{\partial t} + \frac{\partial f}{\partial W_t} m(t, X_t) + \frac{1}{2} \frac{\partial^2 f}{\partial X_t^2} \sigma^2(t, X_t) \right] dt + \frac{\partial f}{\partial W_t} \sigma(t, X_t) dW_t \end{aligned} \quad (57)$$

Processes that satisfy this type of SDE are known as **Itô processes**.

### 3.2.3 Stochastic product rule

We can also formulate the analogue of the product rule  $d(fg) = f'g + fg'$ . Let  $X_t$  and  $Y_t$  be stochastic processes, then

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + d\langle X, Y \rangle_t \quad (58)$$

For example, if  $dX_t = f_1(t)dt + g_1(t)dW_t$  and  $dY_t = f_2(t)dt + g_2(t)dW_t$ , then

$$d(X_t Y_t) = (X_t f_2(t) + Y_t f_1(t) + g_1(t)g_2(t))dt + (g_1 Y_t + g_2 X_t)dW_t \quad (59)$$

### 3.2.4 Examples

- Define the geometric Brownian motion with drift  $\mu$  and volatility  $\sigma$  as

$$dX_t = \mu X_t dt + \sigma X_t dW_t \quad (60)$$

Solving this SDE requires finding a function  $f(t, x)$  satisfying the system of equations

$$\begin{cases} \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} = \mu f \\ \frac{\partial f}{\partial x} = \sigma f \end{cases} \implies f(t, x) = g(t) \exp(\sigma x) \quad (61)$$

The first equation implies that  $g(t)$  should satisfy

$$g'(t) + \frac{1}{2}(\sigma)^2 g(t) = \mu g(t) \implies g(t) = g(0) \exp\left(\mu t - \frac{1}{2}\sigma^2 t\right) \quad (62)$$

Therefore, the closed form of the geometric Brownian motion is

$$X_t = X_0 \exp\left(\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right)t\right) \quad (63)$$

Using the fact that  $\mathbb{E}[e^{\sigma W_t}] = e^{1/2\sigma^2 t}$  we see that  $\mathbb{E}[e^{X_t}] = e^{\mu t}$  so the process is drifting exponentially fast.  $X_t$  is distributed according to a log-normal distribution

$$f(s) = \frac{1}{\sqrt{2\pi(s\sigma)^2 t}} \exp\left(-\frac{(\log s - \log(X_0) - \mu + \frac{1}{2}\sigma^2)^2}{\sigma^2 t}\right) \quad (64)$$

- Consider the exponential SDE

$$dX_t = A_t X_t dW_t \quad (65)$$

To solve this we use Itô's formula consider a generic function  $X_t = f(t, W_t)$

$$df = \left[ \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial W_t^2} \right] dt + \frac{\partial f}{\partial W_t} dW_t \quad (66)$$

Comparing terms by terms we get the sistem of equations

$$\begin{cases} \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial W_t^2} = 0 \\ \frac{\partial f}{\partial W_t} = A_t f \end{cases} \implies f(t, W_t) = g(t) \exp\left(\int_0^t A_s dW_s\right) \quad (67)$$

we then solve for  $g(t)$  using the first equation

$$\frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial W_t^2} = g'(t) + \frac{1}{2}(A_t)^2 g(t) = 0 \implies g(t) = g(0) \exp\left(-\frac{1}{2} \int_0^t ds A_s^2\right) \quad (68)$$

Hence, the solution to the SDE is

$$X_t = X_0 \exp\left(\int_0^t A_s dW_s - \frac{1}{2} \int_0^t ds A_s^2\right) \quad (69)$$

### 3.3 Kolmogorov equations

#### 3.3.1 Markov semigroup

For every Markov process we can define the probability that a process is at  $y$  at time  $t$  given that at time zero it was at  $x$ ; we denote the density as

$$\mathbb{P}[X_t \in dy | X_0 = x] = p(t; x, y) dy \quad (70)$$

We focus on time-homogeneous Markov processes for which the transition probability  $\mathbb{P}((x, t) \rightarrow (x', t'))$  only depends on the time difference (ie.  $\mathbb{P}((x, t) \rightarrow (x', t')) = p(t' - t; x, y)$ ).

The density satisfies the Chapman-Kolmogorov equation

$$p(t; x, y) = \int ds p(t - t'; s, y) p(t'; x, s) \quad (71)$$

which is a consequence of Markov properties. It says that the probability density of going from  $x \rightarrow y$  in time  $t$  is obtained by going from  $x \rightarrow s$  in time  $t'$  and then from  $s$  to  $y$  in the remaining  $t - t'$  and we need to sum over all possible values of  $s$ .

The Chapman-Kolmogorov equation suggest that a *time-homogeneous* Markov process can be described by a semigroup  $P_t$  such that  $P_0 = I$  and  $P_t \circ P_s = P_{t+s} \forall t, s \geq 0$ . We define the action of this semigroup on continous and bounded functions as

$$P_t(f)(x) = \int_{\mathbb{R}^d} f(y)p(t; x, y)dy \quad (72)$$

The Chapman-Kolmogorov equation implies that

$$\begin{aligned} P_{t+s}(f)(x) &= \int_{\mathbb{R}^d} f(y)p(t+s; x, y)dy = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(y)p(t; x, z)p(s; z, y)dydz \\ &= \int_{\mathbb{R}^d} dz P_s(f)(z)p(t; x, z) = (P_t \circ P_s)(f)(x) \end{aligned}$$

Define the generator of the semigroup as an operator  $\mathcal{L}$  such that

$$\mathcal{L}(f) = \lim_{t \rightarrow 0} \frac{P_t(f) - f}{t} \quad (73)$$

this operator is known as the **generator of the Markov process**; furthermore, it satisfies  $P_t = e^{t\mathcal{L}}$ . Consider the quantity  $u(x, t) = P_t(f)(x) = \mathbb{E}[f(X_t)|X_0 = x]$  then it satisfies the differential equation

$$\partial_t u(x, t) = \partial_t e^{t\mathcal{L}} f = \mathcal{L}u(x, t) \quad (74)$$

with initial condition  $u(x, 0) = f(x)$ . This equation is known as **backward Kolmogorov equation**. Notice that we can also define  $\tilde{u}(x, s) = \mathbb{E}[f(X_T)|X_s = x]$  where  $s \in [0, T]$  and  $T$  is fixed. Then using time-homogeneity we have

$$\tilde{u}(x, s) = \mathbb{E}[f(X_T)|X_s = x] = \int_{\mathbb{R}^d} f(y)p(T-s; x, y)dy \quad (75)$$

which implies that  $\partial_T$  is equal to  $-\partial_s$ ; therefore,  $u(x, s)$  satisfies

$$-\partial_s u(x, s) = \mathcal{L}u(x, s) \quad (76)$$

with terminal condition  $u(x, t) = f(x)$ .

Similarly, we can define an operator  $P_t^\dagger$  that acts on probability measure rather than on bounded function, it is the adjoint of  $P_t$  in the  $L^2$  sense.

$$\int_{\mathbb{R}^d} P_t(f)(y)d\mu(y) \equiv \int_{\mathbb{R}^d} f(y)d(P_t^\dagger(\mu)(y)) \quad (77)$$

In a similar way we can define a generator  $\mathcal{L}^\dagger$  such that  $P_t = e^{t\mathcal{L}^\dagger}$  and the generic measure  $\rho(x, t)$  evolves according therefore

$$\partial_t \rho(x, t) = \mathcal{L}^\dagger \rho(x, t) \quad (78)$$

with initial condition  $\rho(x, 0) = \rho_0(x)$ . When the intial condition is deterministic  $X_0 = x$  we have  $\rho(x, 0) = \delta(x)$ . This equation is known as **forward Kolmogorov equation**.

We define a **Ergodic Markov process** if the generator  $\mathcal{L}$  has a zero eigenvalue of algebraic multiplicity 1, i.e.  $\mathcal{L}g = 0$  is solved by  $g = \text{const}$  only. This implies that  $P_t(g) = g$  has only constant solutions  $\forall t \geq 0$  and we can define an invariant measure  $\mu$  satisfying

$$P_t^\dagger \mu = \mu \implies \mathcal{L}^\dagger \rho = 0 \quad d\mu(x) = \rho(x)dx \quad (79)$$

This invariant measure determines the long-time behaviour of a random initial condition  $X_0$  distributed according to  $\mu_0$

$$\lim_{t \rightarrow \infty} P_t^\dagger \mu_0 = \mu \quad (80)$$

we then have the notion of time-average=phase-space average

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t dt' f(X_{t'}) = \int d\mu(x) f(x) \quad (81)$$

### 3.3.2 Diffusion processes

Diffusion processes are special cases of Markov processes for which the following properties hold. Let  $p(y, t|x, s)$  be the probability that  $X_t = y$  given that  $X_s = x$

- Continuity:  $\forall x, \epsilon$

$$\int_{|x-y| \leq \epsilon} dy p(y, t|x, s) = o(t-s) \quad (82)$$

- Drift: there exists  $b$  such that  $\forall x, \epsilon$

$$\int_{|x-y| \leq \epsilon} dy (y-x) p(y, t|x, s) = b(x, s)(t-s) + o(t-s) \quad (83)$$

- Diffusion: there exists  $D$  such that  $\forall x, \epsilon$

$$\int_{|x-y| \leq \epsilon} dy (y-x)^2 p(y, t|x, s) = \Sigma(x, s)(t-s) + o(t-s) \quad (84)$$

Then it can be proved that if  $X_t$  is a diffusion process then

- **Backward Kolmogorov equation:** The quantity  $u(x, s) = \mathbb{E}[f(X_t)|X_s = x]$  satisfies

$$-\partial_s u(x, s) = b(x, s) \partial_x u(x, s) + \frac{1}{2} \Sigma(x, s) \partial_x^2 u(x, s) \quad (85)$$

with terminal condition  $u(x, t) = f(x)$ . Notice that if we let  $f(x) = \delta(x)$ , then we have an evolution equation for the probability  $p(y, t|x, s)$  in terms of the backward variables  $x, s$

$$-\partial_s p(y, t|x, s) = b(x, s) \partial_x p(y, t|x, s) + \frac{1}{2} \Sigma(x, s) \partial_x^2 p(y, t|x, s) \quad (86)$$

- **Forward Kolmogorov equation:** The quantity  $p(y, t|x, s)$  satisfies

$$\partial_t p(y, t|x, s) = -\partial_y [b(y, t) p(y, t|x, s)] + \frac{1}{2} \partial_y^2 [\Sigma(y, t) p(y, t|x, s)] \quad (87)$$

with initial condition  $p(y, s|x, s) = \delta(y-x)$ . Notice that this is in terms of the forward variables  $y, t$  and it is the adjoint of the backward equation (doing integration by parts and pulling derivatives onto the measure).

### 3.3.3 Kramers-Moyal expansion

The Fokker-Plank equation is as special case of Kolmogorov forward equation which we now derive using the Kramers-Moyal expansion.

Consider the quantity, for a generic test function  $h(y)$

$$\begin{aligned}
 \int dy h(y) \partial_t p(t; x, y) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int dy h(y) p(t + \Delta t; x, y) - p(t; x, y) \\
 &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int dy h(y) \left[ \int ds p(\Delta t; s, y) p(t; x, s) - p(t; x, y) \right] \\
 &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int dy ds p(t; x, s) p(\Delta t; s, y) [h(y) - h(s)] \\
 &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int dy ds p(t; x, s) p(\Delta t; s, y) \sum_{n \geq 1} \frac{h^{(n)}(s)}{n!} (y - s)^n
 \end{aligned}$$

Define the jump moments as

$$D^{(n)}(s) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int dy p(\Delta t; s, y) \frac{1}{n!} (y - s)^n \quad (88)$$

we then have

$$\begin{aligned}
 \int dy h(y) \partial_t p(t; x, y) &= \int dy ds p(t; x, s) \sum_{n \geq 1} D^{(n)}(s) h^{(n)}(s) \\
 &= \int dy ds h(s) \sum_{n \geq 1} (-1)^n \partial_s^n [D^{(n)}(s) p(t; x, s)]
 \end{aligned}$$

Using the fact that the test function  $h(y)$  is arbitrary we obtain the **Kolmogorov forward equation**

$$\partial_t p(t; x, y) = \sum_{n \geq 1} (-1)^n \partial_y^n [D^{(n)}(y) p(t; x, y)] \quad (89)$$

if we assume that  $D^{(n)}(y) = 0$  for  $n > 2$  we get the **Fokker-Plank equation**

$$\partial_t p(t; x, y) = -\partial_y [\mu(y) p(t; x, y)] + \frac{1}{2} \partial_y^2 [\sigma^2(y) p(t; x, y)] \quad (90)$$

where we have defined  $D^{(1)}(y) = \mu(y)$  and  $D^{(2)}(y) = 1/2 \sigma^2(y)$ .

For a Itô process

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dW_t$$

the Fokker Plank equation is exact and the generator (backward) in  $d = 1$  is

$$\mathcal{L} = \mu(x, t) \partial_x + \frac{1}{2} \sigma^2(x, t) \partial_x^2 \equiv \mu(x, t) \partial_x + D(x, t) \partial_x^2 \quad (91)$$

where  $D$  is the diffusion constant.

Similarly, in  $d$  dimensions

$$\mathcal{L} = \boldsymbol{\mu}(x, t) \cdot \nabla_x + \frac{1}{2} \text{tr}(\sigma(x, t) \sigma^\top(x, t) H) \equiv \boldsymbol{\mu}(x, t) \cdot \nabla_x + \text{tr}(D(x, t) H) \quad (92)$$

where  $H_{ij} = \partial_{x_i} \partial_{x_j}$  is the Hessian.

We can now see how the diffusion equation

$$\partial_t p(x, t) = D \partial_x^2 p(x, t) \quad (93)$$



for a conserved density  $p(x, t)$  can be associated to a Ito process of the form

$$dX_t = \sqrt{2D}dW_t \quad (94)$$

where  $D$  is the diffusion constant, equivalently we notice that  $dZ_t \equiv dW_{t\sqrt{2D}} = \sqrt{2D}dW_t$  hence we have that the noise is  $dZ_t$  with  $\mathbb{E}[dZ_t dZ_s] = 2D\delta(t-s)$  which is sometimes used in physics.

### 3.4 Stratonovich integration

As we learned, the crucial part of stochastic calculus lies on the interpretation of

$$\int_0^t dW_t f(t, W_t) \quad (95)$$

currently we interpret it in the Itô sense where for a simple process  $A_t$  given by

$$A_t = \sum_{k=1}^n \xi_k \mathbb{1}_{[t_{k-1}, t_k)} \quad (96)$$

we have

$$\int A_s dW_s = \sum_{k=1}^n \xi_i (W_{t_{i+1}} - W_{t_i}) \quad (97)$$

choosing the endpoint  $\xi_i$  is the crucial part of Itô integration. However, we have seen that Itô integrals do not follow the usual rules of calculus. Stratonovich integrals try fix this issue so that

$$\int_0^t f'(W_s) \circ dW_s = f(W_t) - f(W_0) \quad (98)$$

holds.

This is achieved by defining the Stratonovich integral as

$$\int_0^t Y_t \circ dX_t = \int_0^t Y_t dX_t + \frac{1}{2} \langle X, Y \rangle_t \quad (99)$$

This is achieved by using the midpoint  $1/2(\xi_i + \xi_{i+1})$  results in the definition of **Stratonovich integral** as

$$\int A_s \circ dW_s = \sum_{k=1}^n \frac{1}{2} (\xi_i + \xi_{i+1}) (W_{t_{i+1}} - W_{t_i}) \quad (100)$$

$$= \sum_{k=1}^n \xi_i (W_{t_{i+1}} - W_{t_i}) + \sum_{k=1}^n \frac{1}{2} (\xi_{i+1} - \xi_i) (W_{t_{i+1}} - W_{t_i}) \quad (101)$$

$$\rightarrow \int A_s dW_s + \frac{1}{2} \langle A, W \rangle_t \quad (102)$$

for a simple process  $A_t$ . We can see that the extra  $\xi_{i+1}$  will disrupt the proof of this integral begin a martingale. In general, Stratonovich integrals are not martingale. Similarly, the variance rule

$$\mathbb{E}[(\int A_s dW_s)^2] = \int \mathbb{E}[A_s^2] ds \quad (103)$$

does not hold for Stratonovich.

In general, given an Itô process

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t \quad (104)$$

we have an equivalent Stratonovich process

$$dX_t = \left( \mu(X_t, t) - \frac{1}{2} \sigma'(X_t, t) \sigma(X_t, t) \right) dt + \sigma(X_t, t) \circ dW_t \quad (105)$$

where  $\sigma'$  means derivative with respect to the first argument ( $X_t$ ). To prove this we use the definition

$$\int_0^t f(X_s) \circ dW_s = \int_0^t f(X_s) dW_s + \frac{1}{2} \langle f(X), W \rangle_t \quad (106)$$

which in differential form reads

$$f(X_t) \circ dW_t = f(X_t) dW_t + \frac{1}{2} f'(X_t) \sigma dt \quad (107)$$

For the Itô process we have  $f = \sigma$  which yields the previous equation. Notice that we can write the analogue of Itô's lemma

$$\begin{aligned} df(X_t, t) &= f'(X_t, t) dX_t + \frac{1}{2} f''(X_t, t) d\langle X \rangle_t \\ &= f'(X_t, t) \sigma dW_t + \left[ f'(X_t, t) \mu + \frac{1}{2} \sigma^2 f''(X_t, t) \right] dt \\ &= f'(X_t, t) \sigma \circ dW_t + \left[ f'(X_t, t) \mu - \frac{1}{2} (f'' \sigma + f' \sigma') \sigma + \frac{1}{2} \sigma^2 f''(X_t, t) \right] dt \\ &= f'(X_t, t) \sigma \circ dW_t + \left[ f'(X_t, t) \mu - \frac{1}{2} f' \sigma' \sigma \right] dt \end{aligned}$$

Notice that when  $X_t = W_t$  (i.e.  $\sigma = 1$ ,  $\mu = 0$ ) we get the usual rule of differentiation

$$df(W_t) = f'(W_t) \circ dW_t \quad (108)$$

## 4 Important (continuous) processes

This is a section that summarises some results on some important processes

### 4.1 Brownian bridge

Brownian motion, or Wiener process, is a stochastic process that is fixed at  $t = 0$  (usually  $X_0=0$ ) and the correlation function is  $\mathbb{E}[X_t X_s] = \min(s, t)$ . A Brownian is a Brownian motion that is fixed at two ends  $t = 0$  and  $t = T$

$$B_t = \{W_t | W_T = 0\} \quad t \in [0, T] \quad (109)$$

We can construct an analytic expression for  $B_t$

$$B_t = W_t - \frac{t}{T} W_T \quad (110)$$

Notice that

$$\mathbb{E}[B_t] = 0 \quad (111)$$

$$\mathbb{E}[B_t B_s] = \min(t, s) + \frac{ts}{T} - \frac{ts}{T} - \frac{st}{T} = \min(s, t) - \frac{st}{T} \quad (112)$$

Thus the Brownian bridge is Gaussian distributed with

$$B_t \sim \mathcal{N}(0, t(1 - t/T)) \quad (113)$$

Notice that the Brownian bridge can be expressed as

$$B_t = \frac{T-t}{\sqrt{T}} W_{t/(T-t)} \quad (114)$$

because  $\mathbb{E}[B_t] = 0$  and

$$\begin{aligned} \mathbb{E}[B_t B_s] &= \frac{(T-t)(T-s)}{T} \min\left(\frac{t}{T-t}, \frac{s}{T-s}\right) = \min(t(1-s/T), s(1-t/T)) \\ &= \min(s, t) - st/T \end{aligned} \quad (115)$$

Therefore, they  $B_t$  can be re-written as a scaled Brownian motion.

## 4.2 Geometric Brownian motion

Geometric Brownian motion (GBR) is a stochastic process described by the following SDE

$$dS_t = \mu S_t dt + \sigma S_t dW_t \quad (116)$$

which can be solved either as done previously (solving a system of differential equations) or by noticing that

$$d(\ln S_t) = \frac{dS_t}{S_t} - \frac{1}{2} \frac{(\sigma S_t)^2}{S_t^2} dt = \mu dt + \sigma dW_t - \frac{1}{2} \sigma^2 dt \quad (117)$$

which implies

$$S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right) \quad (118)$$

This is used to model asset prices since  $S_t$  is not negative.  $S_t$  is distributed as a log-normal distribution with

$$\mathbb{E}[S_t] = S_0 e^{\mu t} \quad (119)$$

and covariance ( $s < t$ )

$$\begin{aligned} \mathbb{E}[S_t S_s] &= S_0 e^{(\mu - \frac{\sigma^2}{2})(t+s)} \mathbb{E}[e^{\sigma W_t} e^{\sigma W_s}] = S_0 e^{(\mu - \frac{\sigma^2}{2})(t+s)} \mathbb{E}[e^{\sigma(W_t - W_s)} e^{2\sigma W_s}] \\ &= S_0 e^{(\mu - \frac{\sigma^2}{2})(t+s)} \mathbb{E}[e^{\sigma(W_t - W_s)}] \mathbb{E}[e^{2\sigma W_s}] = S_0 e^{(\mu - \frac{\sigma^2}{2})t} e^{\frac{t-s}{2}\sigma^2} e^{2t\sigma^2} \\ &= S_0 e^{(\mu - \frac{\sigma^2}{2})(t+s) + \frac{3t-s}{2}\sigma^2} \end{aligned} \quad (120)$$

In general we can write it as

$$\mathbb{E}[S_t S_s] = S_0 e^{\mu(t+s) + |t-s|\sigma^2} \quad (121)$$

### 4.3 Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck (OU) process satisfies

$$dX_t = -\theta X_t dt + \sigma dW_t \quad (122)$$

the generator (backward) and its adjoint (forward) are given by

$$\mathcal{L} = -\theta x \partial_x + \frac{1}{2} \sigma^2 \partial_x^2 \quad \mathcal{L}^\dagger = \theta \partial_x x + \frac{1}{2} \sigma^2 \partial_x^2 \quad (123)$$

hence the transition probability density satisfies

$$\partial_t p(x, t | x', t') = \theta \partial_x x p(x, t | x', t') + \frac{1}{2} \sigma^2 \partial_x^2 p(x, t | x', t') \quad (124)$$

with  $p(x, t' | x', t') = \delta(x - x')$ .

We first solve for  $X_t$ , the system of equation is given by

$$\begin{cases} \frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial W_t^2} = -\theta f \\ \frac{\partial f}{\partial W_t} = \sigma \end{cases} \implies f(t, W_t) = \sigma W_t + g(t) \quad (125)$$

Plugging into the first equation gives

$$\frac{\partial g}{\partial t} = -\theta(\sigma W_t + g(t)) \quad (126)$$

This is solved by integrating factor multiply both sides by  $h(t)$

$$h(t) \dot{g} = -\theta \sigma W_t h - \theta g h \quad (127)$$

Comparing with  $(\dot{h}g) = \dot{h}g + h\dot{g}$  implies  $\theta gh = \dot{h}g \implies f(t) = e^{\theta t}$ ; hence,  $(\dot{h}g) = -\theta \sigma W_t h$  or  $g(t) = 1/h(t) \int_0^t (-\theta \sigma W_s h(s)) ds = -\theta \sigma e^{-\theta t} \int_0^t W_s e^{\theta s} ds$

The solution of the OU process is

$$X_t = \sigma W_t - \theta \sigma e^{-\theta t} \int_0^t ds e^{\theta s} W_s \quad (128)$$

Notice that if we solve the first equation first, then we have a solution

$$X_t = \sigma e^{-\theta t} \int_0^t dW_s e^{\theta s} \quad (129)$$

these solutions are equivalent since

$$d(e^{\theta t} W_t) = \theta e^{\theta t} W_t dt + e^{\theta t} dW_t \quad (130)$$

Since  $X_t$  is an integral of  $dW_s$  and the integrand does not depend on  $W_t$ ,  $X_t$  is distributed as a Gaussian with zero mean and variance

$$\mathbb{E}[X_t^2] = \sigma^2 e^{-2\theta t} \int_0^t ds e^{2\theta s} = \frac{\sigma^2}{2\theta} (1 - e^{-2\theta t}) \quad (131)$$

Similarly the covariance, or **correlation function**, is given by

$$\begin{aligned} \mathbb{E}[X_t X_s] &= \sigma^2 e^{-\theta(t+s)} \int_0^t dt' \int_0^s ds' e^{\theta(s'+t')} \mathbb{E}[dW_{t'} dW_{s'}] \\ &= \sigma^2 e^{-\theta(t+s)} \int_0^t dt' \int_0^s ds' e^{\theta(s'+t')} \delta(t' - s') = \sigma^2 e^{-\theta(t+s)} \int_0^t dt' e^{2\theta t'} \mathbb{1}_{[0,s]} \\ &= \sigma^2 e^{-\theta(t+s)} \frac{1}{2\theta} (e^{2\theta \min(s,t)} - 1) = \frac{\sigma^2}{2\theta} (e^{-|t-s|} - e^{-\theta(s+t)}) \end{aligned} \quad (132)$$

Notice that when dealing with  $\max(s, t)$  or  $\min(s, t)$  it is usefull sometimes to use the following properties

$$\max(s, t) = \frac{s + t + |s - t|}{2} \quad (133)$$

$$\min(s, t) = \frac{s + t - |s - t|}{2} \quad (134)$$

Notice that we should technically write  $\mathbb{E}[X_t X_s | X_0 = 0]$  i.e we are certain that at  $t = 0$  the process is at the origin. However, we can also have the unconditional covariance where we calculate  $\lim_{s' \rightarrow \infty} \mathbb{E}[X_t X_s | X_{s'} = 0]$  which changes the lower limit of the integral to  $-\infty$  and gives an unconditional correlation function given by

$$C(t, s) = \frac{\sigma^2}{2\theta} e^{-|t-s|} \quad (135)$$

We can write down the transition probability as

$$p(x, t | x', t') = \frac{1}{\sqrt{2\pi \frac{\sigma^2}{2\theta} (1 - e^{-2\theta(t-t')})}} \exp \left[ -\frac{(x - x')^2}{2 \frac{\sigma^2}{2\theta} (1 - e^{-2\theta(t-t')})} \right] \quad (136)$$

which it can be shown to satisfy the appropriate Kolmogorov forward equation (Fokker-Plank). Notice that taking the limit  $|t - t'| \rightarrow \infty$  yields the invariant measure of the OU process

$$p(x) = \frac{1}{\sqrt{2\pi \frac{\sigma^2}{2\theta}}} \exp \left[ -\frac{x^2}{2 \frac{\sigma^2}{2\theta}} \right] \quad (137)$$

which solves the equation  $\mathcal{L}^\dagger p = 0$ .

The OU process is an example of a Gaussian process with a non-vanishing stationary measure.

#### 4.4 Cox-Ingersoll-Ross model

### 5 Path integral

#### 5.1 Feynman-Kac formula

Consider a Itô process and the following expectation value

$$u(x, t) = \mathbb{E}[e^{-\int_0^t V(X_s, s) ds} f(X_t) | X_0 = x] \quad (138)$$

then Feynman-Kac fomula says that  $u(x, t)$  satisfies the following partial differential equation

$$\partial_t u(x, t) = \mu(x, t) \partial_x u(x, t) + \frac{1}{2} \sigma(x, t) \partial_x^2 u(x, t) - V(x, t) u(x, t) \quad (139)$$

with initial condition  $u(x, 0) = f(x)$ .

This can be proved by considering the joint process  $(X_t, Y_t)$  where

$$Y_t = y - \int_0^t V(X_s, s) ds \quad (140)$$

Writing the differential for both processes

$$\begin{aligned} dX_t &= \mu(X_t, t) dt + \sigma(X_t, t) dW_t & X_0 &= x \\ dY_t &= -V(X_t, t) dt & Y_0 &= y \end{aligned}$$

The generator  $\mathcal{L}_{x,y}$  of this process is given by

$$\mathcal{L}_{x,y} = \mu(x,t)\partial_x + \frac{1}{2}\sigma^2(x,t)\partial_x^2 - V(x,t)\partial_y \quad (141)$$

it follows then that  $u(x,y,t) = \mathbb{E}[e^{Y_t}f(X_t)|Y_0 = y, X_0 = x]$  satisfies

$$\partial_t u(x,y,t) = \mathcal{L}_{x,y}u(x,y,t) = \left[ \mu(x,t)\partial_x + \frac{1}{2}\sigma^2(x,t)\partial_x^2 - V(x,t)\partial_y \right] u(x,y,t) \quad (142)$$

with  $u(x,y,0) = e^y f(x)$ . Note that  $u(x,y,t) = u(x,0,t)e^y \equiv u(x,t)e^y$  hence

$$\partial_t u(x,t) = \left[ \mu(x,t)\partial_x + \frac{1}{2}\sigma^2(x,t)\partial_x^2 - V(x,t) \right] u(x,t) \quad (143)$$

It is interesting to note the similarity with a immaginary-time Schrödinger equation (when  $\mu = 0$ ) where  $V$  is just the potential.

Note that the Feynman-Kac formula can also be found using Itô's lemma

$$\begin{aligned} d(e^{-\int_0^t ds V(X_s)} f(X_t)) &= \left[ e^{-\int_0^t ds V(X_s)} \frac{\sigma^2}{2} f''(X_t) + V(X_t) e^{-\int_0^t ds V(X_s)} f(X_t) \right] dt \\ &\quad + e^{-\int_0^t ds V(X_s)} f'(X_t) dX_t \end{aligned} \quad (144)$$

Integrating, taking  $\mathbb{E}$  on both sides and using  $\mathbb{E}[\int_0^t dW_s f(X_s, s)] = 0$ , yields the Feynman-Kac formula.

Similarly, can also calculate the following expectation value

$$u(x,t) = \mathbb{E}[e^{-\int_t^T V(X_s,s)ds} f(X_T)|X_t = x] \quad (145)$$

for  $t \in [0, T]$  where  $T$  is fixed. Using time-homogeneity we can write the correponding PDE just by changing the sign  $\partial_t \rightarrow -\partial_t$

$$-\partial_t u(x,t) = \left[ \mu(x,t)\partial_x + \frac{1}{2}\sigma^2(x,t)\partial_x^2 - V(x,t) \right] u(x,t) \quad (146)$$

## 5.2 Black-Scholes equation

We can derive a famous equation in finance known as Black-Scholes equation which is used to predict the price of an option. An **option** is a financial derivative that gives the holder the right, but not the obligation, to buy or sell an underlying asset at a predetermined price, known as the *strike price*, on or before a specific date, known as the *expiration date*.

There are two main types of options:

- **Call Option:** A call option gives the holder the right to *buy* the underlying asset at the strike price.
- **Put Option:** A put option gives the holder the right to *sell* the underlying asset at the strike price.

How to price an option? In other words, if I tell you the strike price  $K$  at time  $T$  what is a fair price for this option at time  $t$ ?

Let the price of an *option* be denoted a  $V(S_t, t)$  which depends on the underlying asset value  $S_t$ . We assume that  $S_t$  is the price of an asset at time  $t$  which follows a GBM

$$dS_t = \mu S_t dt + \sigma S_t dW_t \quad (147)$$

At the maturation time  $T$  we either sell/buy the option if there is a gain (otherwise we can do not sell/buy). The gain is  $\max(S_T - K, 0)$  or  $\max(K - S_T, 0)$  respectively for call or put options. This is the future value, to get the present value we need to discount it using the continuous compounding hence the present value at time  $t$  is

$$e^{r(T-t)} \max(S_T - K, 0) \quad (148)$$

where  $r$  is a risk-free interest rate (i.e. the return on a risk-free investement e.g. government bonds). This discount reflects the idea that money in the future is worth less than money now. The expectation value of this quantity is the fair price of the call option

$$V(S, t) = \mathbb{E}[e^{r(T-t)} \max(S_T - K, 0) | S_t = S] \quad (149)$$

The Black-Scholes equation is nothing but the application of the Feynman-Kac formula to this expectation which yields

$$-\partial_t V(S, t) = \mu S \partial_S V(S, t) + \frac{1}{2} \sigma^2 S^2 \partial_S^2 V(S, t) - rV(S, t) \quad (150)$$

usually re-written as follows with  $\mu = r$

$$\partial_t V(S, t) + rS \partial_S V(S, t) + \frac{1}{2} \sigma^2 S^2 \partial_S^2 V(S, t) = rV(S, t) \quad (151)$$

## 6 Jump processes

## 7 Fractional Brownian motion