

## Week 2: Stochastic Integrals

Professor Stavros Panageas

Anderson School of Management, MGMTMFE403, Fall 2023

- I. Diffusion and Wiener process
- II. Information, Filtration, Adaptedness
- III. Stochastic Integrals
- IV. Martingales
- V. Quadratic Variation
- VI. Ito's Formula
- VII. Applications of Ito's formula
- VIII. The multi-dimensional Ito's Formula

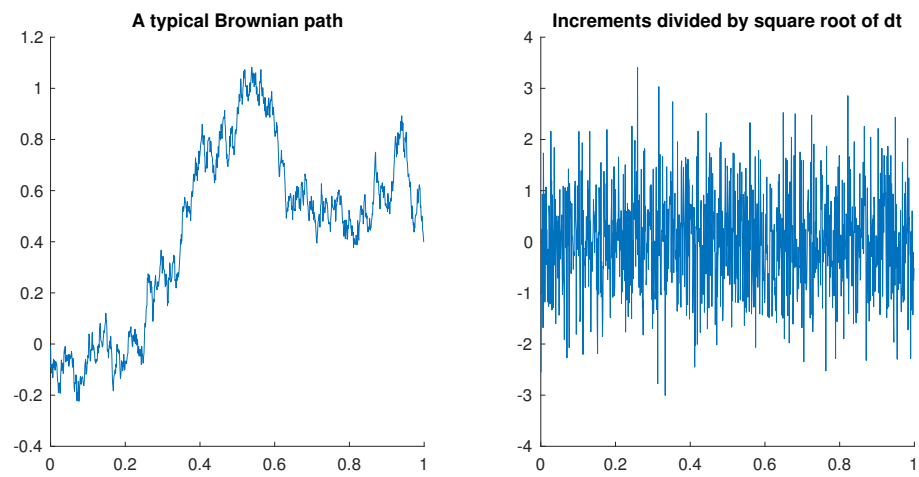
1

### I. Diffusion and Wiener process

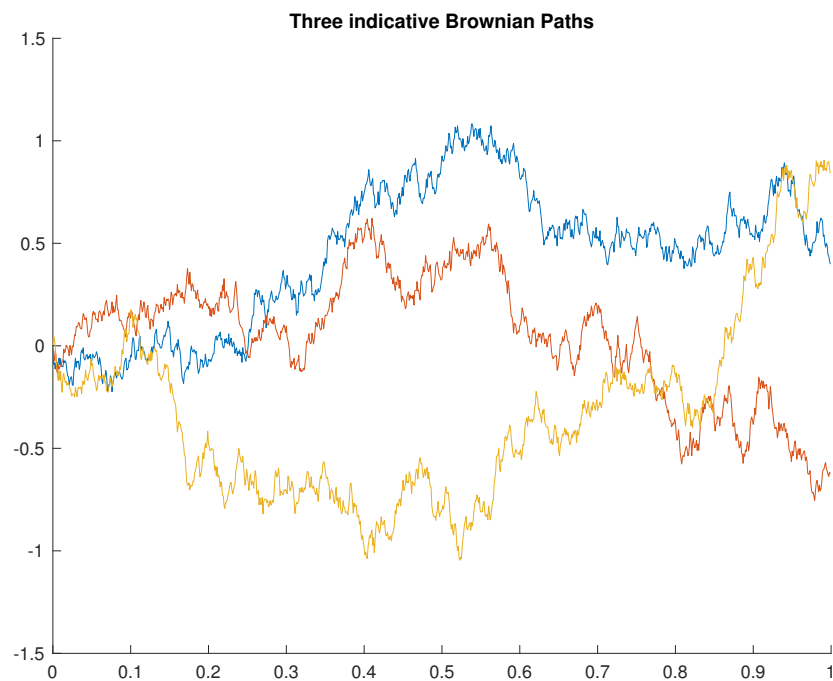
- In this class, we want to generalize the simple binomial dynamics of the previous class.
- Towards that end we introduce the notion of a stochastic process (a function of time that is drawn randomly), which we will refer to as a Wiener process (or Brownian Motion)
- A stochastic process  $W$  is called a **Wiener process** (or equivalently Brownian Motion) if
  1.  $W(0) = 0$
  2.  $W$  has independent increments, i.e., if  $r < s \leq t < u$ , then  $W(u) - W(t)$  and  $W(s) - W(r)$  are independent stochastic variables
  3. For  $s < t$  the stochastic process  $W(t) - W(s)$  has the Gaussian distribution  $N[0, \sqrt{t-s}]$ .
  4.  $W$  has continuous trajectories

2

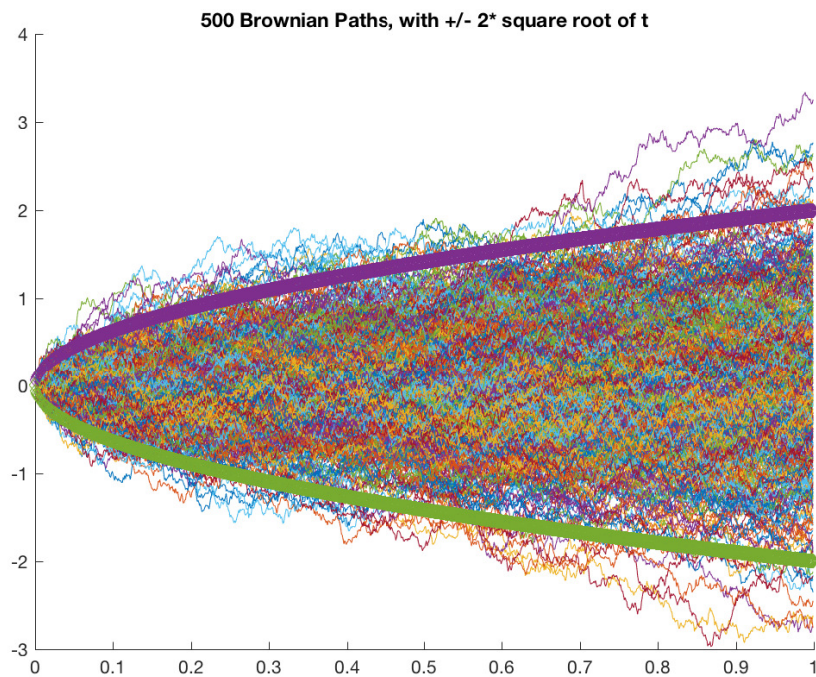
## Illustration of a Wiener process on the time interval $[0, 1]$



3



4



This figure plots 1000 Brownian paths, together with the functions  $2 \times \sqrt{t}$  and  $-2 \times \sqrt{t}$

5

- In many applications we are interested in the dynamics of processes of the form

$$X(t + \Delta t) - X(t) = \mu(t, X(t)) \Delta t + \sigma(t, X(t)) \Delta W(t) \quad (1)$$

where

$$\Delta W(t) = W(t + \Delta t) - W(t)$$

and  $\Delta t$  becomes small.

- The “drift” term  $\mu(t, X(t))$  captures the expected change of  $X(t + \Delta t) - X(t)$  over the next interval  $\Delta t$ .
- The “diffusion” term  $\sigma(t, X(t)) \Delta W(t)$  captures the impact of noise.

- In what sense is there a “solution” to (1)? It looks like a differential equation. So, it is tempting (but terribly wrong, as it turns out) to write it as

$$\dot{X}(t) = \mu(t, X(t)) + \sigma(t, X(t)) u(t),$$

6

where

$$u(t) = \frac{dW_t}{dt}$$

- The next figure shows what is the problem with this idea, namely that the Brownian paths are never differentiable

6

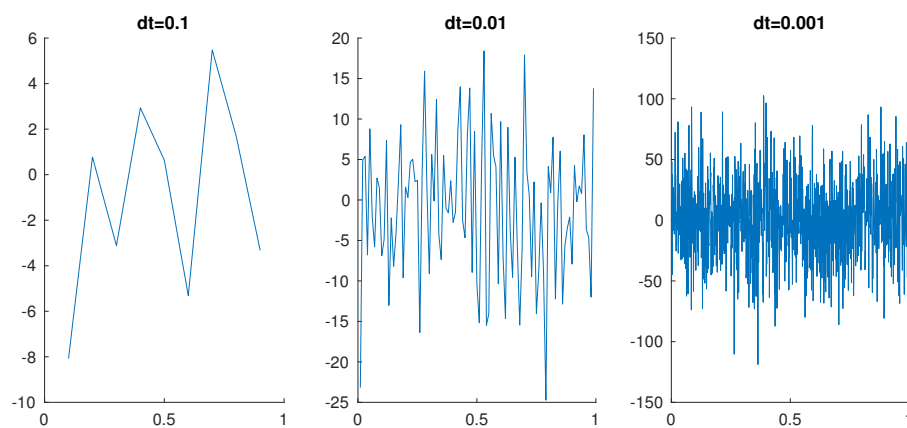


Illustration of the ratio  $\frac{W(t+dt)-W(t)}{dt}$  when  $dt = 0.1$ ,  $dt = 0.01$ ,  $dt = 0.001$ . Note that the paths do not converge to a constant, but instead become increasingly more volatile as  $dt$  shrinks.

7

- Second attempt. How about re-writing the differential equation in integral form

$$X(t) = \alpha + \int_0^t \mu(s, X(s)) ds + \int_0^t \sigma(s, X(s)) dW(s)$$

and interpreting  $\int_0^t \sigma(s, X(s)) dW(s)$  as a Riemann-Stieltjes integral?

- (Aside: A Riemann-Stieltjes integral is defined as an integral against a function such as –say a distribution function–  $\int_0^t \sigma(s, X(s)) dF(s)$ , which may not be differentiable)
- The problem with this idea is that Brownian motion as a function of time:

$$\sup_{P \in \mathcal{P}} \sum_{i=0}^{n_P-1} |W(x_{i+1}) - W(x_i)| = \infty$$

where  $\mathcal{P} = \{P = \{x_0, \dots, x_{n_P}\} | P \text{ is a partition of a finite interval } [a, b]\}$

- Key idea: Let's give up on trying to give a path-wise meaning to " $\int_0^t \sigma(s, X(s)) dW(s)$ ". Instead let's think of it as a probabilistic limit of a sequence of certain random variables.

8

## II. Information, Filtration, Adaptedness

- The symbol  $\mathcal{F}_t^X$  will be used to denote the “information generated by observing a path of the stochastic process  $X$  over the interval  $[0, t]$ ”
- If, based on the observation of the trajectory of  $X_t$  up to the point  $t$  we can determine if a given event  $A$  has occurred, then we will call that event  $\mathcal{F}_t^X$ – “measurable”
- Examples: Are the following events in  $\mathcal{F}_1^X$ ?
  1.  $X_{0.5} > 5$
  2.  $X_{1.5} > 6$
- If  $Y$  is a stochastic process such that we have

$$Y(t) \in \mathcal{F}_t^X$$

then we will call  $Y$  **adapted to the filtration**  $\mathcal{F}_t^X$

- Examples:
  1.  $S(t) = \sup_{s \leq t} W(s)$  is adapted to  $\mathcal{F}_t^X$
  2.  $S(t) = \sup_{s \leq t+1} W(s)$  is not adapted to  $\mathcal{F}_t^X$

9

### III. Stochastic Integrals

- The process  $g$  belongs to the class  $\mathcal{L}^2[a, b]$  if the following conditions are satisfied
  1.  $\int_a^b E[g^2(s)] ds < \infty$
  2. The process  $g$  is adapted to the Filtration  $\mathcal{F}_t^W$
- The process  $g$  belongs to the class  $\mathcal{L}^2$  if it belongs to the class  $\mathcal{L}^2[0, t]$  for all  $t$ .
- A process  $g$  is **simple** if there exist deterministic points in time  $a = t_0 < t_1 < \dots < t_n = b$ , such that  $g$  is constant on that subinterval. In other words  $g(s) = g(t_k)$  for  $s \in [t_k, t_{k+1})$ .
- For simple  $g$ , it is straightforward to define the stochastic integral

$$\int_a^b g(s) dW(s) = \sum_{k=0}^{n-1} g(t_k) [W(t_{k+1}) - W(t_k)]$$

10

- The strategy from now on is to approximate  $g$  with a sequence of simple processes  $g_n$  such that

$$\int_a^b E[\{g_n(s) - g(s)\}^2] ds \rightarrow 0$$

- For each  $n$  the intergral  $\int_a^b g_n(s) dW(s)$  is a well defined stochastic variable
- We are now in a position to define the stochastic integral by

$$\int_a^b g(s) dW(s) = \lim_{n \rightarrow \infty} \int_a^b g_n(s) dW(s)$$

- It can be shown that the limit exists in a probabilistic sense (mean square error convergence).

11

## Properties of Stochastic Integrals

- The following properties hold for the stochastic integral

$$E \int_a^b g(s) dW(s) = 0$$

- The Ito isometry

$$E \left[ \left( \int_a^b g(s) dW(s) \right)^2 \right] = E \int_a^b g^2(s) ds$$

- Let's prove these identities for simple  $g$

12

- Using that fact that  $g$  is simple

$$\begin{aligned} E \left[ \int_a^b g(s) dW(s) \right] &= E \left[ \sum_{k=0}^{n-1} g(t_k) [W(t_{k+1}) - W(t_k)] \right] \\ &= \sum_{k=0}^{n-1} E [g(t_k) E_{t_k} [W(t_{k+1}) - W(t_k)]] \\ &= 0 \end{aligned}$$

- Similarly,

$$\begin{aligned} E \left( \int_a^b g(s) dW(s) \right)^2 &= E \left[ \left( \sum_{k=0}^{n-1} g(t_k) [W(t_{k+1}) - W(t_k)] \right)^2 \right] \\ &= \sum_{k=0}^{n-1} E \left( g^2(t_k) E_{t_k} [W(t_{k+1}) - W(t_k)]^2 \right) \\ &= \sum_{k=0}^{n-1} E (g^2(t_k)) \Delta_{t_k} = \int_a^b E (g^2(s)) ds \end{aligned}$$

13

## IV. Martingales

- A stochastic process  $X$  is called an  $\mathcal{F}_t$  martingale if the following conditions hold

1.  $X$  is adapted to the filtration  $\mathcal{F}_t$  for  $t \geq 0$ .
2. For all  $t$

$$E[|X_t|] < \infty$$

3. For all  $s$  and  $t$  with  $s \leq t$ , the following relation holds

$$E[X(t) | \mathcal{F}_s] = X(s)$$

A process satisfying for all  $s$  and  $t$  with  $s \leq t$  the inequality

$$E[X(t) | \mathcal{F}_s] \leq X(s)$$

is called a supermartingale and a process satisfying the reverse inequality is called a submartingale.

14

- Let  $X_t$  be given by

$$X(t) = \int_0^t g(s) dW(s)$$

- Leaving integrability aside, every stochastic integral is a martingale.
- Proof: Fix any  $u < t$ . We have

$$\begin{aligned} E[X(t) | \mathcal{F}_u] &= E\left[\int_0^t g(s) dW(s) | \mathcal{F}_u\right] \\ &= \int_0^u g(s) dW(s) + E\left[\int_u^t g(s) dW(s) | \mathcal{F}_u\right] \\ &= X(u) + 0 \end{aligned}$$

- Implication: A process  $dX_t$  can only be a martingale if it does not have a “dt” term, that is if

$$dX(t) = g(t) dW(t).$$

15



## V. Quadratic Variation

- In preparation for the development of Ito's formula, we next introduce the notion of "infill asymptotics" by introducing the concept of quadratic variation
- Let us divide the interval  $[0, t]$  into  $n$  equally large subintervals of the form  $\left[k\frac{t}{n}, (k+1)\frac{t}{n}\right]$  where  $k = 0, 1, \dots, n-1$ . Given this subdivision, we now define the quadratic variations of the Wiener process  $S_n$

$$S_n = \sum_{i=1}^n \left[ W\left(i\frac{t}{n}\right) - W\left((i-1)\frac{t}{n}\right) \right]^2$$

- The computation of the expectation of  $S_n$  is easy, and indeed is independent of  $n$  :

$$\begin{aligned} E(S_n) &= \sum_{i=1}^n E \left\{ \left[ W\left(i\frac{t}{n}\right) - W\left((i-1)\frac{t}{n}\right) \right]^2 \right\} \\ &= \sum_{i=1}^n E \left\{ \left( i\frac{t}{n} - (i-1)\frac{t}{n} \right) \right\} = t \end{aligned}$$

16

- Remarkably, the variance of  $S_n$  disappears as  $n \rightarrow \infty$ , as we show next. The fact that  $W_t$  has independent increments implies that

$$\begin{aligned} Var[S_n] &= \sum_{i=1}^n Var \left\{ \left[ W\left(i\frac{t}{n}\right) - W\left((i-1)\frac{t}{n}\right) \right]^2 \right\} \\ &= \sum_{i=1}^n 2 \left[ \frac{t^2}{n^2} \right] = 2\frac{t^2}{n} \end{aligned}$$

- (Hint: To prove that  $Var \left\{ \left[ W\left(i\frac{t}{n}\right) - W\left((i-1)\frac{t}{n}\right) \right]^2 \right\} = 2 \left[ \frac{t^2}{n^2} \right]$  use the fact that  $\left[ W\left(i\frac{t}{n}\right) - W\left((i-1)\frac{t}{n}\right) \right]^2$  is –up to scaling– Chi-square distributed)
- As  $n \rightarrow \infty$ , the variance of  $S_n$  disappears.
- Hence

$$\int_0^t [dW]^2 = t$$

which is sometimes written in short-hand form as

$$[dW]^2 = dt$$

17

## VI. Ito's formula

- Assume that the process  $X$  is given by

$$dX(t) = \mu(t) dt + \sigma(t) dW(t)$$

- where  $\mu$  and  $\sigma$  are adapted processes and let  $f$  be a  $C^{1,2}$  function. Define the process  $Z$  by  $Z(t) = f(t, X(t))$ . Then  $Z$  has a stochastic differential given by

$$df(t, X(t)) = \left\{ \frac{\partial f}{\partial t} + \mu(t) \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2(t) \frac{\partial^2 f}{\partial x^2} \right\} dt + \sigma(t) \frac{\partial f}{\partial x} dW(t)$$

- Let's start with a heuristic proof. Using Taylor's theorem, we obtain

$$df = \left\{ \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dX \right\} + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (dX)^2 + \frac{1}{2} \frac{\partial^2 f}{\partial t^2} (dt)^2 + \frac{\partial^2 f}{\partial x \partial t} dX dt$$

- Heuristically, we can write

$$(dX)^2 = \mu^2(dt)^2 + 2\mu\sigma(dt)dW + \sigma^2(dW)^2$$

where we have used the short-hand notation  $\mu, \sigma$  for  $\mu(t), \sigma(t)$ .

18

- The terms of order  $(dt)^2$  are tiny compared to  $dt$  as  $dt \rightarrow 0$ , so we can ignore them.
- It can also be shown that the  $dt dW$ – term is negligible compared to the  $dt$  term (in a sense we will make more precise shortly). Finally we have  $(dW)^2 = dt$  and hence plugging back into the expression for  $df$  we get

$$df = \left\{ \frac{\partial f}{\partial t} + \mu(t) \frac{\partial f}{\partial x} + \frac{1}{2} \sigma^2(t) \frac{\partial^2 f}{\partial x^2} \right\} dt + \sigma(t) \frac{\partial f}{\partial x} dW(t) + o(dt).$$

- It is sometimes more convenient to remember the Ito Formula as

$$df = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dX + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (dX)^2$$

and use the following formal multiplication rules

$$\begin{aligned} (dt)^2 &= 0 \\ dt dW &= 0 \\ (dW)^2 &= dt \end{aligned}$$

19

## VII. Applications of Ito's formula - Example 1

- Compute the moment generating function of brownian motion  $E \left[ e^{aW(t)} \right]$
- Let  $Z(t) = e^{aW(t)}$
- Apply Ito's formula to compute

$$dZ(t) = aZ(t) dW(t) + \frac{a^2}{2} Z(t) dt, \quad Z(0) = 1$$

- Write this in integral form

$$Z(t) = 1 + \frac{a^2}{2} \int_0^t Z(s) ds + a \int_0^t Z(s) dW(s) \quad (2)$$

- Let  $m(t) = E \left[ e^{aW(t)} \right]$ . Then taking expectations on both sides of (2) leads to

$$m(t) = 1 + \frac{a^2}{2} \int_0^t m(s) ds.$$

20

- or in differential form

$$\dot{m}(t) = \frac{a^2}{2} m(t), \quad m(0) = 1$$

Hence

$$m(t) = e^{\frac{a^2}{2}t}$$

21

## Example 2

- Compute the distribution of

$$\int_0^t W(s) dW(s)$$

- Let  $Z(t) = W^2(t)$ . Applying Ito's Lemma gives

$$dZ(t) = dt + 2W(t) dW(t)$$

Integrating gives

$$W^2(t) = \int_0^t dZ(s) = t + 2 \int_0^t W(s) dW(s)$$

- Therefore

$$\int_0^t W(s) dW(s) = \frac{W^2(t)}{2} - \frac{t}{2}$$

- Therefore (up to scaling by  $\frac{1}{2}$  and translation by  $\frac{t}{2}$ ) the stochastic integral  $\int_0^t W(s) dW(s)$  is distributed as a (scaled) chi-squared variable

22

## Example 3

- Compute the distribution of

$$dX(t) = \alpha X(t) dt + \sigma X(t) dW(t), \quad X(0) = 1$$

- Let  $Z(t) = \log(X_t)$ . Applying Ito's Lemma gives

$$\begin{aligned} dZ(t) &= \frac{1}{X_t} (\alpha X(t) dt + \sigma X(t) dW(t)) - \frac{\sigma^2 X_t^2}{2 X_t^2} dt \\ &= \left( \alpha - \frac{\sigma^2}{2} \right) dt + \sigma dW(t). \end{aligned}$$

Integrating gives

$$\log(X_t) = \left( \alpha - \frac{\sigma^2}{2} \right) t + \sigma W(t)$$

- Hence,

$$X_t = \exp \left\{ \left( \alpha - \frac{\sigma^2}{2} \right) t + \sigma W(t) \right\}$$

is log-normally distributed with mean  $e^{\alpha t}$ .

23

## VIII. The multi-dimensional Ito's Formula

- Take a vector  $X = (X_1, \dots, X_n)$ , where the component  $X_i$  has a stochastic differential of the form

$$dX_i(t) = \mu_i(t) dt + \sum_{j=1}^d \sigma_{ij}(t) dW_j(t)$$

and  $W_1 \dots W_d$  are  $d$  independent Wiener processes.

- Defining the drift vector  $\mu$  by

$$\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix},$$

the  $d$ -dimensional vector Wiener process  $W$  by

$$W = \begin{bmatrix} W_1 \\ \vdots \\ W_d \end{bmatrix}$$

24

and the  $n \times d$  dimensional diffusion matrix by

$$\sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \dots & \sigma_{nd} \end{bmatrix}.$$

We can then write the dynamics of the  $X(t)$  process as

$$dX(t) = \mu(t) dt + \sigma(t) dW(t)$$

Now letting  $f : R_+ \times R^n \rightarrow R$  be  $C^{1,2}$ . Then

$$df(t, X(t)) = \frac{\partial f}{\partial t} dt + \sum_{i=1}^n \frac{\partial f}{\partial x_i} dX_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} dX_i dX_j$$

24

using the formal multiplication rules

$$\begin{aligned}(dt)^2 &= 0 \\ dtdW &= 0 \\ (dW_i)^2 &= dt, \quad i = 1, \dots, d \\ dW_i dW_j &= 0, \quad \text{for } i \neq j\end{aligned}$$

24

### Example 4

- Suppose that

$$\begin{aligned}dX_1 &= \mu_1 X_1 dt + \sigma_{11} X_1 dW_1(t) + \sigma_{12} X_1 dW_2(t) \\ dX_2 &= \mu_2 X_2 dt + \sigma_{22} X_2 dW_2(t)\end{aligned}$$

- Compute

$$d(X_1(t) \cdot X_2(t))$$

- Solution: Let  $Z(t) = X_1(t) \cdot X_2(t)$ . Applying Ito's Lemma we have

$$\begin{aligned}dZ(t) &= Z(t) [(\mu_1 + \mu_2) dt + \sigma_{11} dW_1(t) + (\sigma_{12} + \sigma_{22}) dW_2(t)] \\ &\quad + Z(t) \sigma_{12} \sigma_{22} dt\end{aligned}$$

25

## Example 5

- Suppose that

$$\begin{aligned}dX_1 &= \mu_1 X_1 dt + \sigma_{11} X_1 dW_1(t) + \sigma_{12} X_1 dW_2(t) \\dX_2 &= \mu_2 X_2 dt + \sigma_{22} X_2 dW_2(t)\end{aligned}$$

- Compute

$$d\left(\frac{X_1(t)}{X_2(t)}\right)$$

- Solution: Let  $Z(t) = \frac{X_1(t)}{X_2(t)}$ . Then –applying Ito's Lemma– we have

$$\begin{aligned}dZ(t) &= Z(t) [(\mu_1 - \mu_2) dt + \sigma_{11} dW_1(t) + (\sigma_{12} - \sigma_{22}) dW_2(t)] \\&\quad - Z(t) \sigma_{12} \sigma_{22} dt + Z(t) \sigma_{22}^2 dt\end{aligned}$$

26

## Appendix: A more rigorous derivation of Ito's Formula

- Let's start by writing

$$f(t, X(t)) - f(0, X(0)) = \sum_{k=0}^{n-1} f(t_{k+1}, X(t_{k+1})) - f(t_k, X(t_k)).$$

Taylor's theorem implies that

$$f(t_{k+1}, X(t_{k+1})) - f(t_k, X(t_k)) = f_t \Delta t + f_x \Delta X_k + \frac{1}{2} f_{xx} (\Delta X_k)^2 + Q_k$$

where  $Q_k$  is a remainder term and  $f_t = \frac{\partial f(t_k, X_{t_k})}{\partial t}$ ,  $f_x = \frac{\partial f(t_k, X_{t_k})}{\partial X}$ , etc.

- Moreover, we have that

$$\begin{aligned}\Delta X_k &= X(t_{k+1}) - X(t_k) = \int_{t_k}^{t_{k+1}} \mu(s) ds + \int_{t_k}^{t_{k+1}} \sigma(s) dW(s) \\&= \mu(t_k) \Delta t + \sigma(t_k) \Delta W_{t_k} + S_{t_k}\end{aligned}$$

27

- Hence

$$(\Delta X_k)^2 = \mu^2(t_k) (\Delta t)^2 + 2\mu(t_k) \sigma(t_k) \Delta t \Delta W_{t_k} + \sigma^2(t_k) (\Delta W_{t_k})^2 + P_k,$$

where  $P_k$  is a remainder term.

- Collecting all the above equations into one gives

$$f(t, X(t)) - f(0, X(0)) = I_1 + I_2 + I_3 + \frac{1}{2}I_4 + \frac{1}{2}K_1 + K_2 + R$$

where

$$\begin{aligned} I_1 &= \sum_k f_t \Delta t, \quad I_2 = \sum_k f_x(t_k) \mu(t_k) \Delta t, \quad I_3 = \sum_k f_x \sigma(t_k) \Delta W_k \\ I_4 &= \sum_k f_{xx} \sigma^2(t_k) (\Delta W_k)^2, \quad K_1 = \sum_k f_{xx}(t_k) \mu^2(t_k) (\Delta t)^2, \\ K_2 &= \sum_k f_{xx}(t_k) \mu(t_k) \sigma(t_k) (\Delta t) (\Delta W_k), \\ R &= \sum_k \{Q_k + S_k + P_k\}, \end{aligned}$$

28

- Letting  $n \rightarrow \infty$ , we have by definition

$$\begin{aligned} I_1 &\rightarrow \int_0^t f_t(s, X(s)) ds, \quad I_2 \rightarrow \int_0^t f_x(s, X(s)) \mu(s) ds, \\ I_3 &\rightarrow \int_0^t f_x(s, X(s)) \sigma(s) dW(s) \end{aligned}$$

- Using essentially the same proof as the one we used earlier to show that  $\sum_k (\Delta W_k)^2 \rightarrow t$ , we can show that

$$I_4 \rightarrow \int_0^t f_{xx}(s, X(s)) \sigma^2(s) ds$$

- $K_1$  converges to  $(\Delta t) \times \left( \int_0^t \mu^2(s) ds \right) \rightarrow 0$ , and  $K_2$  has expectation 0 and variance that goes to zero as  $\Delta t \rightarrow 0$ :

$$\begin{aligned} \text{Var}(K_2) &= \sum_k f_{xx}^2(t_k) \mu^2(t_k) \sigma^2(t_k) (\Delta t)^2 \text{Var}(\Delta W_k) \\ &= \sum_k f_{xx}^2(t_k) \mu^2(t_k) \sigma^2(t_k) (\Delta t)^3 \rightarrow 0 \end{aligned}$$

29



- The more difficult part of the proof is to show that the term  $R$  goes to zero, which we omit here. However, the fact that the sums  $I_1 - I_4$  go to well-defined limits, while the sums  $K_1$  and  $K_2$  vanish as  $\Delta t$  goes to zero shows why we use the formal multiplication rules

$$\begin{aligned}(dt)^2 &= 0 \\ dtdW &= 0 \\ (dW)^2 &= dt\end{aligned}$$