

# ML HW10

**Date:** 11/05/2025

**Week:** 10

**Author:** Alvin B. Lin

**Student ID:** 112652040

## Discussion 1: Itô's Lemma

Before answering the problem, I would like to prove Itô's lemma my own. Unlike the proof during the course, I would like to show it with a perspective of **variations**.

### Definition 1: Variation

For **continuous** processes  $\phi, \psi : \mathbb{R} \rightarrow \mathbb{R}$ , we define the  $d$ -variation of  $\phi$  on  $[0, T]$  by:

$$[\phi]_d := \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} |\phi(t_{i+1}) - \phi(t_i)|^d$$

And co-variation is defined as:

$$[\phi, \psi]_d := \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} |\phi(t_{i+1}) - \phi(t_i)| \cdot |\psi(t_{i+1}) - \psi(t_i)|,$$

where for all  $n \in \mathbb{Z}^+$ ,  $0 = t_0 \leq t_1 \leq \dots \leq t_n = T$ ,  $\Pi$  is the partition.

### Corollary 1

A continuous process has finite variation must have **zero** as its quadratic variation.

*Proof.*

$$0 \leq [\phi]_2 = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} |\phi(t_{i+1}) - \phi(t_i)|^2 \leq \limsup_{\|\Pi\| \rightarrow 0} |\phi(t_{i+1}) - \phi(t_i)| \cdot [\phi]_1 = 0 \cdot L = 0$$

Hence  $[\phi]_2 = 0$  is proved. □

### Corollary 2

A continuous process and a finite variation process have **zero** as their co-variation.

*Proof.* Let  $\phi$  be a finite variation process, and  $\psi$  be a continuous process, so

$$0 \leq [\phi, \psi] = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} |\psi(t_{i+1}) - \psi(t_i)| \cdot |\phi(t_{i+1}) - \phi(t_i)| \leq \limsup_{\|\Pi\| \rightarrow 0} |\psi(t_{i+1}) - \psi(t_i)| \cdot [\phi]_1 = 0$$

Hence  $[\phi, \psi] = 0$  is proved □

Now we consider variations on two main process:

- **Identity map:**  $\mathbf{id} : [0, T] \rightarrow \mathbb{R}; t \mapsto t$

$$[\mathbf{id}]_1 = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} |\mathbf{id}(t_{i+1}) - \mathbf{id}(t_i)| = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} |t_{i+1} - t_i| = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} (t_{i+1} - t_i) = T$$

Hence

$$[\mathbf{id}]_1 = T$$

- **Wiener process:**  $W_t$  on  $[0, T]$ .

For an equal length partition  $\Pi$  on  $[0, T]$ , we have

$$[W_t]_1 = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}| = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} \left| \sqrt{\frac{T}{n}} \varepsilon \right|, \quad \varepsilon \sim \mathcal{N}(0, 1)$$

Since  $|\varepsilon|$  is a standard [half-normal distribution](#), meaning that:

$$\mathbb{E}|\varepsilon| = 2 \int_0^{+\infty} \frac{1}{\sqrt{2\pi}} z e^{-\frac{1}{2}z^2} dz = -\frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \Big|_0^{+\infty} = \sqrt{\frac{2}{\pi}}$$

By taking expected value on  $[W_t]$ , we obtain:

$$\mathbb{E}([W_t]) = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} \sqrt{\frac{T}{n}} \mathbb{E}|\varepsilon| = \lim_{n \rightarrow \infty} \left( n \cdot \sqrt{\frac{T}{n}} \cdot \sqrt{\frac{2}{\pi}} \right) = \sqrt{\frac{2T}{\pi}} \lim_{n \rightarrow \infty} \sqrt{n} = +\infty$$

Hence we conclude that

$$[W_t]_1 = +\infty$$

Also, by corollary 1 and 2, we have:

$$[\mathbf{id}]_2 = 0 \quad \text{and} \quad [\mathbf{id}, W_t] = 0$$

#### Lemma 1: Variation on SDE

For time  $t$  and the Browning motion/Wiener process  $W_t$ , we have:

$$(dt)^2 = 0; \quad dt dW_t = 0; \quad (dt)^m (dW_t)^n = 0 \quad \text{for } m, n \in \mathbb{Z}^+$$

*Proof.* By definition,

$$(dt)^2 \equiv d[\mathbf{id}, \mathbf{id}] = d[\mathbf{id}]_2 = d(0) = 0$$

and

$$dt dW_t \equiv d[\mathbf{id}, W_t] = d(0) = 0.$$

Also apply double induction, we can get:

$$(dt)^m (dW_t)^n = 0 \quad \text{for } m, n \in \mathbb{Z}^+$$

□

### Lemma 2: Quadratic Variation of Wiener Process

For Wiener Process  $W_t$  defined on  $[0, T]$ , we have:

$$(dW_t)^2 = dt$$

*Proof.* Recall that for  $0 \leq s \leq t \leq T$ , we have:

$$W_t - W_s \sim \mathcal{N}(0, t - s) \quad \text{or} \quad W_t - W_s \sim \sqrt{t - s}\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, 1)$$

So, we take the expected value on quadratic variation on  $W_t$ , we get:

$$\mathbb{E}([W_t]_2) = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} \mathbb{E}[(\sqrt{t_{i+1} - t_i}\varepsilon)^2] = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} (t_{i+1} - t_i) \mathbb{E}(\varepsilon^2) \quad (1)$$

Also, by definition of variance:

$$1 = \text{Var}(\varepsilon) = \mathbb{E}(\varepsilon^2) - [\mathbb{E}(\varepsilon)]^2 = \mathbb{E}(\varepsilon^2) - 0^2 \implies \mathbb{E}(\varepsilon^2) = 1 \quad (2)$$

With equation (1) and (2), we obtain:

$$\mathbb{E}([W_t]_2) = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} (t_{i+1} - t_i) = T$$

Consider the variance of the quadratic variation on  $W_t$ , we have:

$$\text{Var}([W_t]_2) = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} \text{Var}((t_{i+1} - t_i)\varepsilon^2) = \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} (t_{i+1} - t_i)^2 \text{Var}(\varepsilon^2)$$

We will use 4th moment and 2nd moment to solve for  $\text{Var}(\varepsilon^2)$ :

$$\text{Var}(\varepsilon^2) = \mathbb{E}(\varepsilon^4) - (\mathbb{E}(\varepsilon^2))^2 = 3 - 1^2 = 2$$

Therefore

$$\text{Var}([W_t]_2) = 2 \lim_{\|\Pi\| \rightarrow 0} \sum_{i=0}^{n-1} (t_{i+1} - t_i)^2 = 2[\mathbf{id}]_2 = 0$$

Zero variance implies the random variable is a constant, hence

$$(dW_t)^2 = d[W_t]_2 = dt$$

□

The two lemmas have shown the calculation law in Itô's calculus, one difference is that the general chain rule **does not hold** in Itô's calculus.

$$d\phi(x, t) \neq \frac{\partial \phi}{\partial x} dx + \frac{\partial \phi}{\partial t} dt$$

### Theorem 1: Itô's Lemma

for a 1-dimension Itô process  $\phi(x, t) \in \mathcal{C}^2(\mathbb{R} \times \mathbb{R})$ , we have:

$$d\phi = \frac{\partial \phi}{\partial t} dt + \frac{\partial \phi}{\partial x} dx + \frac{1}{2} \frac{\partial^2 \phi}{\partial x^2} (dx)^2$$

More generally, for an Itô process  $\Phi(X, t) \in \mathcal{C}^2(\mathcal{X} \times \mathbb{R})$ ,  $\mathcal{X} \subseteq \mathbb{R}^d$ , we have:

$$d\Phi = \frac{\partial \Phi}{\partial t} dt + \nabla \Phi \cdot dX + \frac{1}{2} \sum_{i=1}^d \sum_{k=1}^d \frac{\partial^2 \Phi}{\partial X^i \partial X^k} dX^i dX^k$$

*Proof.* By **Taylor theorem** and lemma 1 and 2,

$$\begin{aligned} d\phi &= \frac{\partial \phi}{\partial x} dx + \frac{\partial \phi}{\partial t} dt + \frac{1}{2} \frac{\partial^2 \phi}{\partial x^2} (dx)^2 + \frac{1}{2} \frac{\partial^2 \phi}{\partial t^2} (dt)^2 + \frac{\partial^2 \phi}{\partial t \partial x} dt dx + \dots \\ &= \frac{\partial \phi}{\partial x} dx + \frac{\partial \phi}{\partial t} dt + \frac{1}{2} \frac{\partial^2 \phi}{\partial x^2} (dx)^2 + 0 \end{aligned}$$

□

Then we've done the proof.

### Discussion 2: Fokker-Planck Equation

Another thing we need to solve for the problem is **Fokker-Planck equation**. Similar to the setting in the resource[1] and the lecture note[2], we consider  $p(x, t) \in \mathcal{C}^2(\mathbb{R} \times \mathbb{R})$  and let

$$dx_t = f(x, t)dt + g(x, t)dW_t$$

For **arbitrary** function with compact support  $V(x) \in \mathcal{C}_c^\infty(\mathbb{R})$ , with Itô's lemma, we have

$$\begin{aligned} dV(x) &= \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial x} dx + \frac{1}{2} \frac{\partial^2 V}{\partial x^2} (dx)^2 = 0 + \frac{\partial V}{\partial x} (f(x, t)dt + g(x, t)dW_t) + \frac{1}{2} \frac{\partial^2 V}{\partial x^2} g^2(x, t)dt \\ &= \left( \frac{\partial V}{\partial x} f(x, t) + \frac{1}{2} \frac{\partial^2 V}{\partial x^2} g^2(x, t) \right) dt + \frac{\partial V}{\partial x} g(x, t)dW_t \end{aligned}$$

### Lemma 3

For predictable (without  $W_t$ ),  $\mathbf{L}^2$ -integrable process  $H_t$ , we must have

$$\mathbb{E} \left( \int H_t dW_t \right) = 0$$

*Proof.* Sketch: use simple function to approximate  $H_t$ , as [Here](#)

□

Hence taking the expected value at the both sides in the above equation, resulting in

$$\mathbb{E}(V(x)) = \mathbb{E} \left( \int dV \right) = \mathbb{E} \left( \int \frac{\partial V}{\partial x} f(x, t) + \frac{1}{2} \frac{\partial^2 V}{\partial x^2} g^2(x, t) dt \right) + \mathbb{E} \left( \int \cancel{\frac{\partial V}{\partial x} g(x, t) dW_t} \right) \rightarrow 0$$

Then take the derivative with respect to  $t$ ,

$$\begin{aligned}
\frac{d}{dt} (\mathbb{E}(V(x))) &= \frac{d}{dt} \left( \mathbb{E} \left( \int \frac{\partial V}{\partial x} f(x, t) + \frac{1}{2} \frac{\partial^2 V}{\partial x^2} g^2(x, t) dt \right) \right) \\
&= \mathbb{E} \left( \frac{d}{dt} \left( \int \frac{\partial V}{\partial x} f(x, t) + \frac{1}{2} \frac{\partial^2 V}{\partial x^2} g^2(x, t) dt \right) \right) \\
&= \mathbb{E} \left( \frac{\partial V}{\partial x} f(x, t) + \frac{1}{2} \frac{\partial^2 V}{\partial x^2} g^2(x, t) \right) \\
&= \int_{\mathbb{R}} \left( \frac{\partial V}{\partial x} f(x, t) + \frac{1}{2} \frac{\partial^2 V}{\partial x^2} g^2(x, t) \right) p(x, t) dx
\end{aligned}$$

Recall the original definition of the expected value, which is:

$$\mathbb{E}(V(x)) = \int_{\mathbb{R}} V(x) p(x, t) dx$$

Taking the derivative with respect to  $t$  in this expression yields:

$$\frac{d}{dt} (\mathbb{E}(V(x))) = \int_{\mathbb{R}} V(x) \frac{\partial p(x, t)}{\partial t} dx =$$

Two expressions must be equivalent, so:

$$\begin{aligned}
\int_{\mathbb{R}} V(x) \frac{\partial p(x, t)}{\partial t} dx &= \int_{\mathbb{R}} \left( \frac{\partial V}{\partial x} f(x, t) + \frac{1}{2} \frac{\partial^2 V}{\partial x^2} g^2(x, t) \right) p(x, t) dx \\
&= \underbrace{\int_{\mathbb{R}} \frac{\partial V}{\partial x} f(x, t) p(x, t) dx}_{(i)} + \frac{1}{2} \underbrace{\int_{\mathbb{R}} \frac{\partial^2 V}{\partial x^2} g^2(x, t) p(x, t) dx}_{(ii)}
\end{aligned}$$

Remember that  $V(x)$  has compact support, so

$$\lim_{x \rightarrow \pm\infty} V(x) = 0; \quad \lim_{x \rightarrow \pm\infty} \frac{\partial V}{\partial x} = 0$$

We aim to factor out all the  $V(x)$ , so we do the integration by parts on (i) and (ii).

(i)

$$\int_{\mathbb{R}} \frac{\partial V}{\partial x} f(x, t) p(x, t) dx = \cancel{V(x) f(x, t) p(x, t) \Big|_{-\infty}^{+\infty}}^0 - \int_{\mathbb{R}} V(x) \frac{\partial (f(x, t) p(x, t))}{\partial x} dx$$

(ii)

$$\begin{aligned}
\int_{\mathbb{R}} \frac{\partial^2 V}{\partial x^2} g^2(x, t) p(x, t) dx &= \cancel{\frac{\partial V}{\partial x} g^2(x, t) p(x, t) \Big|_{-\infty}^{+\infty}}^0 - \int_{\mathbb{R}} \frac{\partial V}{\partial x} \frac{\partial (g^2(x, t) p(x, t))}{\partial x} dx \\
&= -V(x) \frac{\partial (g^2(x, t) p(x, t))}{\partial x} \Big|_{-\infty}^{\infty} + \int_{\mathbb{R}} V \frac{\partial^2 (g(x, t) p(x, t))}{\partial x^2} dx
\end{aligned}$$

Bring back the result in (i) and (ii),

$$\int_{\mathbb{R}} V(x) \frac{\partial p(x, t)}{\partial t} dx = \int_{\mathbb{R}} V(x) \left( -\frac{\partial(f(x, t)p(x, t))}{\partial x} + \frac{1}{2} \frac{\partial^2(g^2(x, t)p(x, t))}{\partial x^2} \right) dx$$

Since our function  $V(x)$  is **arbitrary**, so

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial(f(x, t)p(x, t))}{\partial x} + \frac{1}{2} \frac{\partial^2(g^2(x, t)p(x, t))}{\partial x^2}, \quad (1)$$

which is the **Fokker Planck equation**.

### Problem 1: PF-ODE Derivation

1. Consider a forward SDE

$$dx_t = f(x_t, t) dt + g(x_t, t) dW_t,$$

show that the corresponding probability flow ODE is written as

$$dx_t = \left( f(x_t, t) - \frac{1}{2} \frac{\partial}{\partial x} g^2(x_t, t) - \frac{g^2(x_t, t)}{2} \frac{\partial}{\partial x} \log p(x_t, t) \right) dt.$$

**Solution:** Before solving the equation, we need another physics theorem.

#### Theorem 2: Continuity Equation (Physics)

For the density  $\rho(x, t)$  of a conserved quantity at position time  $x$  and  $t$ ,  $J(x, t)$  be the flux, we have the following relation:

$$\frac{\partial \rho(x, t)}{\partial t} = -\frac{\partial J(x, t)}{\partial x}$$

An intuitive example to the theorem is

“Imagine a single line of people moving into a stadium through a narrow turnstile.”

- State Space ( $x$ ): The position along the line.
- Density ( $\rho$ ): The number of people per meter in the line at position  $x$ .
- Flux ( $J$ ): The number of people passing a point  $x$  per second.

**Does our  $p(x, t)$  fit in the equation?**

**Yes**, because  $p(x, t)$  is conservative  $\left( \int_{\mathbb{R}} p(x, t) dx = 1 \quad \forall t \right)$

So let's assume some **speed**  $u(x, t)$  s.t.  $u(x, t)p(x, t) = J(x, t)$  and satisfy the equation:

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial J(x, t)}{\partial x} = -\frac{\partial(u(x, t)p(x, t))}{\partial x} \quad (2)$$

We bring equation (2) to equation (1) to solve for  $u(x, t)$ :

$$-\frac{\partial(u(x, t)p(x, t))}{\partial x} = \frac{\partial p(x, t)}{\partial t} = -\frac{\partial(f(x, t)p(x, t))}{\partial x} + \frac{\partial}{\partial x} \left( \frac{1}{2} \frac{\partial(g^2(x, t)p(x, t))}{\partial x} \right) \quad (3)$$

Integrate the equation (3) with respect to  $x$ , we get:

$$\begin{aligned} -u(x, t)p(x, t) &= -f(x, t)p(x, t) + \frac{1}{2} \frac{\partial(g^2(x, t)p(x, t))}{\partial x} \\ &= -f(x, t)p(x, t) + \frac{1}{2} \frac{\partial g^2(x, t)}{\partial x} p(x, t) + \frac{1}{2} g^2(x, t) \frac{\partial p(x, t)}{\partial x} \\ &= -f(x, t)p(x, t) + \frac{1}{2} \frac{\partial g^2(x, t)}{\partial x} p(x, t) + \frac{1}{2} g^2(x, t) \frac{\partial \log p(x, t)}{\partial x} p(x, t) \end{aligned} \quad (4)$$

We cancel out  $p(x, t)$  at both sides, getting:

$$u(x, t) = f(x, t) - \frac{1}{2} \frac{\partial g^2(x, t)}{\partial x} - \frac{g^2(x, t)}{2} \frac{\partial \log p(x, t)}{\partial x}$$

Look back the definition of  $u(x, t)$  from the equation (2), we know that  $u(x, t)$  stands for the speed of flow, which can also be defined as  $\frac{dx_t}{dt}$ , so

$$dx_t = u(x_t, t)dt = \left( f(x_t, t) - \frac{1}{2} \frac{\partial g^2(x_t, t)}{\partial x} - \frac{g^2(x_t, t)}{2} \frac{\partial \log p(x_t, t)}{\partial x} \right) dt,$$

which completes the demonstration.

## References

- [1] Ji-Ha Kim. Deriving reverse-time stochastic differential equations (sdes). <https://jiha-kim.github.io/posts/deriving-reverse-time-stochastic-differential-equations-sdes/#a1-sketch-of-the-fokkerplanck-derivation>, 2025. Accessed: 7 November 2025.
- [2] Tesheng Lin. 2025\_ml\_week\_9. [https://hackmd.io/@teshenglin/2025\\_ML\\_week\\_9](https://hackmd.io/@teshenglin/2025_ML_week_9), 2025. Accessed: 7 November 2025.