MS&E 322: Lecture Notes for Itô Formula

Wenhao Yang

Management Science and Engineering, Stanford University

October 13, 2025

Roadmap

In this lecture, our goal is to figure out how to represent the stochastic process $f(X_t)$ with a general function $f(\cdot)$. In calculus, if X_t is a deterministic and differentiable function w.r.t. time t, Newton-Leibniz formula tells us:

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dX_s.$$

However, in stochastic calculus, directly applying the Newton-Leibniz formula is incorrect. A simple example is $f(x) = x^2$ and $X_t = B_t$:

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)dB_s + 2t = f(B_0) + 2\int_0^t B_s dB_s + 2t.$$

Thus, we need a new formula to extend the Newton-Leibniz formula to fit in stochastic calculus, which is called Itô formula. We will study Itô formula in the following line:

(1) 1 dimension and $X_t = B_t$.

(2) 1 dimension and X_t is an Itô process.

(3) Multi-dimension.

In this course, we restrict our analysis on the following stochastic process class.

Definition 0.1 (Itô Process)**.** *For a stochastic process $X(\cdot)$ on $(\Omega, \mathbb{P}, \mathcal{F})$, we call it Itô process if there exists $F \in \mathcal{L}^1$ and $G \in \mathcal{L}^2$ such that:*

$$X(\cdot) = X(0) + \int_0^{\cdot} F(s)ds + \int_0^{\cdot} G(s)dB_s.$$

We also denote the process in differential form:

$$dX_t = F(t)dt + G(t)dB_t.$$

1 Warm-up: Itô Formula with $X_t = B_t$

Let's do an intuitive analysis, which is mathematically incorrect. By previous lectures, we know:

$$dB_t \approx \sqrt{dt} > dt,$$

which means dB_t is a half-order term dominating the first order term dt. Thus, by applying Taylor expansion, we have:

$$\begin{aligned} df(B_t) &\approx f(B_0) + f'(B_t)dB_t + \frac{1}{2}f''(B_t)(dB_t)^2 + o((dB_t)^2) \\ &\approx f(B_0) + f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt + o(dt). \end{aligned}$$

Thus, the Itô formula should be:

Theorem 1.1 (Itô formula 1)**.** *Suppose $f(\cdot) \in C^2$, then we have:*

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds.$$

Equivalently, we also denote it in differentiable form:

$$df(B_t) = f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt.$$

In the following, we prove Theorem 1.1 rigorously. We consider a partition $\Pi_n : 0 = t_0 < t_1 < \cdots < t_n = t$ and $\Delta_n = \max_{i=1,\cdots n} |t_i - t_{i-1}|$, we have:

$$\begin{aligned} f(B_t) - f(B_0) &= \sum_{i=1}^n \left(f(B_{t_i}) - f(B_{t_{i-1}})\right) \\ &\overset{(a)}{=} \sum_{i=1}^n f'(\xi_{t_i})\left(B_{t_i} - B_{t_{i-1}}\right), \text{ where } \xi_{t_i} \in [B_{t_i}, B_{t_{i-1}}] \text{ or } \xi_{t_i} \in [B_{t_{i-1}}, B_{t_i}] \\ &= \sum_{i=1}^n f'(B_{t_{i-1}})\left(B_{t_i} - B_{t_{i-1}}\right) + \sum_{i=1}^n \left(f'(\xi_{t_i}) - f'(B_{t_{i-1}})\right)\left(B_{t_i} - B_{t_{i-1}}\right) \end{aligned}$$

where (a) is due to mean value theorem. For the first term, by definition of stochastic integral, it satisfies:

$$\sum_{i=1}^n f'(B_{t_{i-1}})\left(B_{t_i} - B_{t_{i-1}}\right) \overset{L_2}{\to} \int_0^t f'(B_s)dB_s. \tag{1}$$

For the second term, we notice that, in the worst scenario that $f'(\cdot)$ is Lipchitz, we have:

$$\begin{aligned} &\sum_{i=1}^n \left(f'(\xi_{t_i}) - f'(B_{t_{i-1}})\right)\left(B_{t_i} - B_{t_{i-1}}\right) \\ =&\frac{1}{2}\sum_{i=1}^n f''(\eta_{t_i})\left(B_{t_i} - B_{t_{i-1}}\right)^2, \text{ where } \eta_{t_i} \in [B_{t_i}, B_{t_{i-1}}] \text{ or } \eta_{t_i} \in [B_{t_{i-1}}, B_{t_i}]. \end{aligned}$$

$$=\frac{1}{2}\sum_{i=1}^{n} f''(B_{t_{i-1}})\left(B_{t_i}-B_{t_{i-1}}\right)^2+\frac{1}{2}\sum_{i=1}^{n}\left(f''(\eta_{t_i})-f''(B_{t_{i-1}})\right)\left(B_{t_i}-B_{t_{i-1}}\right)^2.$$

We notice that $\left(B_{t_i}-B_{t_{i-1}}\right)^2 \approx t_i - t_{i-1}$ and $\left|f''(\eta_{t_i})-f''(B_{t_{i-1}})\right| \approx 0$, where proof is deferred to Lemma 1.1 and Lemma 1.2. Then, we would like to expect the second term would converge to:

$$\sum_{i=1}^{n}\left(f'(\xi_{t_i})-f'(B_{t_{i-1}})\right)\left(B_{t_i}-B_{t_{i-1}}\right) \to \frac{1}{2}\int_0^t f''(B_s)ds. \tag{2}$$

Combining Eqn (1) and Eqn (2), we have:

$$f(B_t)-f(B_0)=\int_0^t f'(B_s)dB_s+\frac{1}{2}\int_0^t f''(B_s)ds.$$

Lemma 1.1. *For a bounded and continuous function $g(\cdot)$, we have:*

$$\sum_{i=1}^{n} g(B_{t_{i-1}})\left(B_{t_i}-B_{t_{i-1}}\right)^2 \xrightarrow{L_2} \int_0^t g(B_s)ds.$$

Lemma 1.2. *For a bounded and uniform continuous function $g(\cdot)$ and $\{t_i\}_{i=0}^{2^n}$ is an equal partition on $[0,t]$ with size 2^{-n}, for any $\eta_{t_i} \in [B_{t_{i-1}}, B_{t_i}]$, we have:*

$$\sum_{i=1}^{2^n}\left(g(\eta_{t_i})-g(B_{t_{i-1}})\right)\left(B_{t_i}-B_{t_{i-1}}\right)^2 \xrightarrow{a.s.} 0.$$

Corollary 1.1. *Suppose $f(t,x) \in C^1 \times C^2$, then we have:*

$$f(t,B_t)=f(0,B_0)+\int_0^t \frac{\partial f(s,B_s)}{\partial s}ds+\int_0^t \frac{\partial f(s,B_s)}{\partial x}dB_s+\frac{1}{2}\int_0^t \frac{\partial^2 f(s,B_s)}{\partial x^2}ds.$$

Equivalently, we also denote it in differentiable form:

$$df(t,B_t)=f_t'(t,B_t)dt+f_x'(t,B_t)dB_t+\frac{1}{2}f_{xx}''(t,B_t)dt.$$

Example 1.1. $\exp\left(cB_t-\frac{c^2}{2}t\right)$ *is a martingale w.r.t.* $(\mathcal{F}_t)_{t\geq 0}$.

Proof. Let $f(t,x)=\exp\left(cx-\frac{c^2}{2}t\right)$, we have:

$$f_t'(t,x)=\exp\left(cx-\frac{c^2}{2}t\right)\cdot\left(-\frac{c^2}{2}\right),$$
$$f_x'(t,x)=\exp\left(cx-\frac{c^2}{2}t\right)\cdot c,$$
$$f_{xx}''(t,x)=\exp\left(cx-\frac{c^2}{2}t\right)\cdot c^2.$$

Apply Corollary 1.1, we have:

$$df(t,B_t)=f(t,B_t)dB_t.$$

Thus, it is martingale. □

2 Itô Formula with Itô process X_t

Similarly, we also do an intuitive analysis at first. By definition 3.1, we have:

$$dX_t = F(t)dt + G(t)dB_t.$$

Then, by Taylor expansion, we have:

$$\begin{aligned}
df(X_t) &\approx f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2 \\
&\approx f'(X_t)F(t)dt + f'(X_t)G(t)dB_t + \frac{1}{2}f''(X_t)\left(F(t)^2(dt)^2 + 2F(t)G(t)dtdB_t + G(t)^2(dB_t)^2\right) \\
&= \left(f'(X_t)F(t) + \frac{1}{2}f''(X_t)G(t)^2\right)dt + f'(X_t)G(t)dB_t + o(dt)
\end{aligned}$$

The rigorous statement is given as follows.

Theorem 2.1 (Itô formula 2)**.** *Suppose X_t is an Itô process*

$$dX_t = F(t)dt + G(t)dB_t,$$

and $f \in C^2(\mathbb{R})$. Then

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)F(s)ds + \int_0^t f'(X_s)G(s)dB_s + \frac{1}{2}\int_0^t f''(X_s)G(s)^2 ds.$$

Equivalently, in differential form:

$$df(X_t) = \left(f'(X_t)F(t) + \frac{1}{2}f''(X_t)G(t)^2\right)dt + f'(X_t)G(t)dB_t.$$

Corollary 2.1. *Suppose X_t is an Itô process*

$$dX_t = F(t)dt + G(t)dB_t,$$

and $f \in C^1(\mathbb{R}) \times C^2(\mathbb{R})$. Then

$$\begin{aligned}
f(t, X_t) =& f(0, X_0) + \int_0^t \frac{\partial f(s, X_s)}{\partial s}ds + \int_0^t \frac{\partial f(s, X_s)}{\partial x}F(s)ds + \frac{1}{2}\int_0^t \frac{\partial^2 f(s, X_s)}{\partial x^2}G(s)^2 ds. \\
&+ \int_0^t \frac{\partial f(s, X_s)}{\partial x}G(s)dB_s
\end{aligned}$$

Equivalently, in differential form:

$$df(t, X_t) = \left(f_t'(t, X_t) + f_x'(t, X_t)F(t) + \frac{1}{2}f_{xx}''(t, X_t)G(t)^2\right)dt + f_x'(t, X_t)G(t)dB_t.$$

Example 2.1. *Suppose ξ_t is a bounded stochastic process adaptive to $(\mathcal{F}_t)_{t\geq 0}$, then*

$$\exp\left(\int_0^t \xi_s dB_s - \frac{1}{2}\int_0^t \xi_s^2 ds\right)$$

is a martingale w.r.t. $(\mathcal{F}_t)_{t\geq 0}$.

Proof. Let $f(x) = e^x$ and $X_t = \int_0^t \xi_s dB_s - \frac{1}{2}\int_0^t \xi_s^2 ds$. Applying Itô formula, we have:

$$\begin{aligned} df(X_t) &= \left(e^{X_t}\left(-\frac{1}{2}\xi_t^2\right) + \frac{1}{2}e^{X_t}\xi_t^2\right)dt + e^{X_t}\xi_t dB_t \\ &= f(X_t)\xi_t dB_t. \end{aligned}$$

Thus, $f(X_t)$ is a martingale. □

3 Multi-dimensional Itô Formula

We denote the d-dimensional Brownian motion as $\mathbf{B}_t = (B_t^{(1)}, \cdots, B_t^{(d)})$, where coordinates are jointly independent. Similarly, we define the multi-dimensional Itô process in the following.

Definition 3.1 (Multi-dimensional Itô Process)**.** *For a d-dimension stochastic process* $\mathbf{X}(\cdot)$*, we call it Itô process if there exists* $\mathbf{F}^{(i)} \in \mathcal{L}^1$ *and* $\mathbf{G}^{(ij)} \in \mathcal{L}^2$ *such that:*

$$\mathbf{X}(\cdot) = \mathbf{X}(0) + \int_0^{\cdot} \mathbf{F}(s)ds + \int_0^{\cdot} \mathbf{G}(s)d\mathbf{B}_s.$$

We also denote the process in differential form:

$$d\mathbf{X}_t = \mathbf{F}(t)dt + \mathbf{G}(t)d\mathbf{B}_t.$$

We do an intuitive analysis by Taylor expansion:

$$\begin{aligned} df(t, \mathbf{X}_t) &= f_t'(t, \mathbf{X}_t)dt + \nabla_x f(t, \mathbf{X}_t)^\top d\mathbf{X}_t + \frac{1}{2}d\mathbf{X}_t^\top \nabla_{xx}^2 f(t, \mathbf{X}_t)d\mathbf{X}_t \\ &= f_t'(t, \mathbf{X}_t)dt + \nabla_x f(t, \mathbf{X}_t)^\top d\mathbf{X}_t + \frac{1}{2}d\mathbf{B}_t^\top \mathbf{G}(t)^\top \nabla_{xx}^2 f(t, \mathbf{X}_t)\mathbf{G}(t)d\mathbf{B}_t \end{aligned}$$

By the independence for different coordinates, we have $dB_t^{(i)} dB_t^{(j)} = \delta_{ij}dt$, which implies:

$$df(t, \mathbf{X}_t) = f_t'(t, \mathbf{X}_t)dt + \nabla_x f(t, \mathbf{X}_t)^\top d\mathbf{X}_t + \frac{1}{2}\operatorname{Tr}\left(\mathbf{G}(t)^\top \nabla_{xx}^2 f(t, \mathbf{X}_t)\mathbf{G}(t)\right)dt$$

Lemma 3.1. *For two independent Brownian motion* $B_t^{(1)}$ *and* $B_t^{(2)}$*, we consider a partition* Π_n : $0 = t_0 < t_1 < \cdots < t_n = t$ *and denote:*

$$S_n = \sum_{i=1}^n \left(B_{t_i}^{(1)} - B_{t_{i-1}}^{(1)}\right)\left(B_{t_i}^{(2)} - B_{t_{i-1}}^{(2)}\right).$$

Then, we have $\mathbb{E}[S_n] = 0$ *and* $\mathbb{E}[S_n^2] \to 0$.

Theorem 3.1 (Itô formula 3)**.** *Suppose* $\mathbf{X}_t$ *is a d-dimension Itô process*

$$d\mathbf{X}_t = \mathbf{F}(t)dt + \mathbf{G}(t)d\mathbf{B}_t,$$

and $f \in C^1(\mathbb{R}) \times C^2(\mathbb{R}^d)$*. Then*

$$df(t, \mathbf{X}_t) = \left(f_t'(t, \mathbf{X}_t) + \nabla_x f(t, \mathbf{X}_t)^\top \mathbf{F}(t) + \frac{1}{2}\operatorname{Tr}\left(\mathbf{G}(t)^\top \nabla_{xx}^2 f(t, \mathbf{X}_t)\mathbf{G}(t)\right)\right)dt + f'(\mathbf{X}_t)\mathbf{G}(t)d\mathbf{B}_t.$$

By the multi-dimensional Itô formula, we can obtain the partial integration theorem by setting $f(x, y) = xy$.

Theorem 3.2. *Suppose X_t and Y_t are two Itô process, then we have:*

$$d(X_tY_t) = X_tdY_t + Y_tdX_t + dX_tdY_t.$$

4 Applications

Example 4.1 (Ruin Problem)**.** *Suppose $X_t = X_0 + bt + \sigma B_t$ is a drifted Brownian motion, $X_0 = x$, and $x \in (l, r)$. We denote the hitting time $\tau_l = \inf\{t|X_t = l\}$ and $\tau_r = \inf\{t|X_t = r\}$. What is the probability X_t reaches l first, i.e. $\mathbb{P}_x(\tau_l < \tau_r)$. What is the expectation, i.e. $\mathbb{E}[\tau_l]$.*

Proof. We consider constructing a martingale $f(X_t)$ and applying optional sampling theorem. By Itô formula, we have:

$$df(X_t) = \sigma f'(X_t)dB_t + \left(bf'(X_t) + \frac{1}{2}\sigma^2 f''(X_t)\right) dt.$$

Thus, we only need to solve the following ODE to make $f(X_t)$ be a martingale:

$$-\frac{2b}{\sigma^2}f'(x) = f''(x).$$

It is obvious that $f(x) = \exp\left(-\frac{2b}{\sigma^2}x\right)$. Furthermore, we denote $\tau = \inf\{t|X_t \notin (l, r)\}$. By optional sampling theorem (verify the conditions!) and have:

$$\mathbb{E}[f(X_\tau)] = \mathbb{E}[f(X_0)] = \exp\left(-\frac{2b}{\sigma^2}x\right).$$

On the other side, we have:

$$\begin{aligned}\mathbb{E}[f(X_\tau)] &= \mathbb{E}[f(X_r)|\tau = \tau_r]\mathbb{P}(\tau = \tau_r) + \mathbb{E}[f(X_l)|\tau = \tau_l]\mathbb{P}(\tau = \tau_l)\\ &= f(r)\mathbb{P}(\tau_r < \tau_l) + f(l)\mathbb{P}(\tau_r > \tau_l).\end{aligned}$$

Thus, we conclude:

$$\mathbb{P}(\tau_r > \tau_l) = \frac{f(r) - f(x)}{f(r) - f(l)}.$$

For expectation $\mathbb{E}[\tau_l]$, we consider constructing a martingale $g(t, X_t) = h(X_t) + t$ and applying optional sampling theorem. By Itô formula, we have:

$$dg(t, X_t) = \sigma h'(X_t)dB_t + \left(1 + bh'(X_t) + \frac{\sigma^2}{2}f''(X_t)\right) dt.$$

Thus, we only need to solve the followinng ODE to make $g(t, X_t)$ be a martingale:

$$1 + bh'(X_t) + \frac{\sigma^2}{2}f''(X_t) = 0,$$

which is:

$$h(x) = A + B\exp\left(-\frac{2b}{\sigma^2}x\right) - \frac{x}{b}.$$

Furthermore, we require $g(0, l) = g(0, r) = 0$ and $= 0$:

$$A + B\exp\left(-\frac{2b}{\sigma^2}l\right) - \frac{l}{b} = 0,$$
$$A + B\exp\left(-\frac{2b}{\sigma^2}r\right) - \frac{r}{b} = 0.$$

We obtain:

$$A = \frac{\frac{r}{b}\exp\left(-\frac{2b}{\sigma^2}l\right) - \frac{l}{b}\exp\left(-\frac{2b}{\sigma^2}r\right)}{\exp\left(-\frac{2b}{\sigma^2}l\right) - \exp\left(-\frac{2b}{\sigma^2}r\right)},$$
$$B = \frac{\frac{l-r}{b}}{\exp\left(-\frac{2b}{\sigma^2}l\right) - \exp\left(-\frac{2b}{\sigma^2}r\right)}.$$

Then, by optional sampling theorem, we have:

$$\mathbb{E}[g(\tau, X_\tau)] = \mathbb{E}[g(0, X_0)] = h(x).$$

Besides, we notice:

$$\mathbb{E}[g(\tau, X_\tau)] = \mathbb{E}[g(0, X_\tau)] + \mathbb{E}[\tau] = \mathbb{E}[\tau].$$

Thus, we have:

$$\mathbb{E}[\tau] = \frac{l-x}{b} + \frac{l-r}{b}\frac{\exp\left(-\frac{2b}{\sigma^2}x\right) - \exp\left(-\frac{2b}{\sigma^2}l\right)}{\exp\left(-\frac{2b}{\sigma^2}l\right) - \exp\left(-\frac{2b}{\sigma^2}r\right)}.$$

When $b < 0$, we let $r \to +\infty$ and have:

$$\mathbb{E}[\tau_l] = \frac{l-x}{b}.$$

When $b > 0$, we let $r \to +\infty$ and have:

$$\mathbb{E}[\tau_l] = +\infty.$$

For $b = 0$, we construct a new martingale $Z(X_t) = X_t^2 - \sigma^2 t^2$. Proof is similar. $\square$