[AI2613 Lecture 13] Diffusion, Itô Integral, Itô Formula

May 31, 2022

1 Itô Integral

Recall that in the last lecture, we formalized a diffusion $\{X(t)\}$ as

$$\mathrm{d}X(t) = \mu(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t),$$

where $\{W(t)\}$ is a standard Brownian motion[1]. With this formalization, the motion of $\{X(t)\}$ in a tiny time interval $[t, t+h]$ can be view as a moving under the differential equation $\frac{\mathrm{d}X(t)}{\mathrm{d}t} = \mu(t, X(t))\,\mathrm{d}t$ with a random perturbation $\sigma(t, X(t))\,\mathrm{d}W(t)$. In this lecture, we define the mathematical meaning of the above stochastic differential equation.

[1] Here we generalize the specification in the last lecture and allow μ and σ^2 to be functions of two variables, time t and position $X(t)$.

Given an ordinary differential equation $\mathrm{d}f(t) = f(t)\,\mathrm{d}t$, we have that,

$$\forall T,\ \int_0^T \mathrm{d}f(t) = \int_0^T f(t)\,\mathrm{d}t,$$

which is equivalent to

$$\forall T,\ f(t) = f(0) + \int_0^T f(t)\,\mathrm{d}t.$$

If we apply the same process to the stochastic differential equation

$$\mathrm{d}X(t) = \mu(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t),$$

we have that

$$\forall T,\ X(T) = X(0) + \int_0^T \mu(t, X(t))\,\mathrm{d}t + \int_0^T \sigma(t, X(t))\,\mathrm{d}W(t). \tag{1}$$

If we regard $\{X(t)\}$ as a function defined on the sample space Ω, for a fixed $\omega \in \Omega$, $\{X_\omega(t)\}$ and $\{W_\omega(t)\}$ are fixed paths. Then Equation (1) means $\forall T, \forall \omega \in \Omega$,

$$X_\omega(T) = X_\omega(0) + \int_0^T \mu(t, X_\omega(t))\,\mathrm{d}t + \int_0^T \sigma(t, X_\omega(t))\,\mathrm{d}W_\omega(t).$$

Note that $\int_0^T \mu(t, X_\omega(t))\,\mathrm{d}t$ is the ordinary Riemann integral of $\mu(t, X_\omega(t))$. The main goal today is to rigorously define the meaning of $\int_0^T \sigma(t, X_\omega(t))\,\mathrm{d}W_\omega(t)$. Our first try is the Riemann-Stieltjes integral.

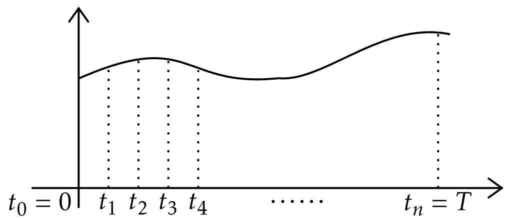
1.1 Riemann-Stieltjes Integral

Recall that when we define the Riemann integral of function g on $[0, T]$, we divide the interval into n segments $[t_0, t_1], [t_1, t_2], \ldots, [t_{n-1}, t_n]$

where each $t_i - t_{i-1} \to 0$ when $n \to \infty$. Then the Riemann integral is defined by

$$\int_0^T g(t)\,\mathrm{d}t = \lim_{n\to\infty} \sum_{i=1}^n g(t_i^*)(t_i - t_{i-1}),$$

where t_i^* can be an arbitrary point in $[t_{i-1}, t_i]$.



Let $F\colon [0,T] \to \mathbb{R}$ be a nice enough function. Assume notations defined above. Then the Riemann-Stieltjes integral of g with respect to F is defined by

$$\int_0^T g(t)\,\mathrm{d}F(t) \triangleq \lim_{n\to\infty} \sum_{i=1}^n g(t_i^*)(F(t_i) - F(t_{i-1})).$$

The Riemann-Stieltjes integral is commonly used in probability theory. Let X be a random variable on sample space $[0,1]$. Assume that the CDF of X is F and the PDF of X is f. Then the expectation of X is $\mathbf{E}[X] = \int_0^1 X\,dF(t)$. By the definition of the Riemann-Stieltjes integral, we have

$$\int_0^1 X\,dF = \lim_{n\to\infty} X(t_i^*)(F(t_i) - F(t_{i-1})) = \lim_{n\to\infty} X(t_i^*) f(t_{i-1})(t_i - t_{i-1}) + o(1),$$

which yields that $\mathbf{E}[X] = \int_0^1 X(t)\,dF(t) = \int_0^T X(t) f(t)\,dt$.

When F is the standard Brownian motion $\{W(t)\}$, the Riemann-Stieltjes integral indicates that

$$\int_0^T g(t)\,\mathrm{d}W(t) = \lim_{n\to\infty} \sum_{i=1}^n g(t_i^*)(W(t_i) - W(t_{i-1})).$$

However, since the Brownian motion is not a function of bounded variation, $\sum_{i=1}^n g(t_i^*)(W(t_i) - W(t_{i-1}))$ may not be convergent[2]. Thus, we cannot find a random variable Y such that for almost every $\omega \in \Omega$, $Y(\omega) = \lim_{n\to\infty} \sum_{i=1}^n g(t_i^*)(W_\omega(t_i) - W_\omega(t_{i-1}))$. This indicates that $\int_0^T g(t)\,\mathrm{d}W(t)$ is not well-defined with the Riemann-Stieltjes integral.

[2] It might be convergent for some specific functions g. However, this does not hold in general.

1.2 *Itô Integral*

Consider the example of $\int_0^T W(t)\,\mathrm{d}W(t)$. Let $\Delta_i W \triangleq W(t_i) - W(t_{i-1})$ and $S_n \triangleq \sum_{i=1}^n W(t_{i-1})\Delta_i W$. By direct calculation, we have that

$$S_n = \frac{1}{2} W_T^2 - \frac{1}{2} \sum_{i=1}^n (\Delta_i W)^2.$$

Let $Q_n \triangleq \sum_{i=1}^n (\Delta_i W)^2$. Note that $\Delta_i W \sim \mathcal{N}(0, \Delta_i)$ where $\Delta_i \triangleq t_i - t_{i-1}$. Then

$$\mathbf{E}[Q_n] = \sum_{i=1}^n \mathbf{E}\left[(\Delta_i W)^2\right] = \sum_{i=1}^n \Delta_i = T,$$

and

$$\begin{aligned}\mathbf{Var}[Q_n] &= \sum_{i=1}^n \mathbf{Var}\left[(\Delta_i W)^2\right] = \sum_{i=1}^n \mathbf{E}\left[(\Delta_i W)^4\right] - \sum_{i=1}^n \mathbf{E}\left[(\Delta_i W)^2\right]^2 \\ &= 2\sum_{i=1}^n \Delta_i^2 \le 2\left(\max_{i\in[n]} \Delta_i\right)\cdot \sum_{i=1}^n \Delta_i \overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

Note that $\mathbf{Var}[Q_n] = \mathbf{E}\left[(Q_n - \mathbf{E}[Q_n])^2\right] = \mathbf{E}\left[(Q_n - T)^2\right]$, and therefore $\mathbf{E}\left[(Q_n - T)^2\right] \to 0$ as $n \to \infty$. This means that Q_n converges to T is the following mean square sense.

Definition 1 (Mean Square Convergence) *Let $Z_1, Z_2, \ldots$ and Z be random variables that $\mathbf{E}\left[Z^2\right] < \infty$ and $\mathbf{E}\left[Z_n^2\right] < \infty$ for $n \ge 1$. We say $Z_n \to Z$ in mean square, or Z is the mean square limit of $\{Z_n\}$, if* $\lim_{n\to\infty} \mathbf{E}\left[(Z_n - Z)^2\right] = 0$.

Itô integral is defined in a similar way.

Definition 2 *Assume that $\{X(t)\}$ is a "nice enough" stochastic process* [3]*. Then we define the integral $\int_0^T X(t)\,\mathrm{d}W(t)$ as the mean square limit of*

$$\sum_{i=1}^n X(t_{i-1})(W(t_i) - W(t_{i-1})).$$

This is called the Itô integral of $\{X(t)\}$ with respect to $\{W(t)\}$.

[3] This means that the process has some nice properties such as $X(t)$ is independent with $\{W(u)\}_{u>t}$ for all t. Some other technical requirements can be found in any standard textbook on stochastic differential equations (e.g., [Kle12])

With Definition 2, we can verify that

$$\int_0^T W(t)\,\mathrm{d}W(t) = \frac{1}{2}W_T^2 - \frac{1}{2}T.$$

More generally, we may define $\int_0^T X(t)\,\mathrm{d}W(t)$ as the mean square limit of

$$\sum_{i=1}^n X(t_i^*)(W(t_i) - W(t_{i-1}))$$

where $t_i^* = \alpha t_{i-1} + (1-\alpha)t_i$ with $\alpha \in [0,1]$. The Itô integral defined in Definition 2 corresponds to the case that $\alpha = 1$. By choosing $\alpha = \frac{1}{2}$, we have the definition of Stratonovich integral and it holds that $\int_0^T W(t)\,\mathrm{d}W(t) = \frac{1}{2}W_T^2$ with Stratonovich integral.

Since we can define integrals for all $\alpha \in [0,1]$, it is natural to ask which value of α produces the "correct" one? Actually, the best choice of α depends on how you want to model the stochastic process. For example, when we view a diffusion $\{X_t\}$ as the limit of a certain discrete process, the motion during $[t, t+h]$ for a tiny h is specified by $\mu(t, X_t)$ and $\sigma^2(t, X_t)$. So it is reasonable to specify a diffusion with Itô integral. However, for many stochastic processes from physics which are continuous in nature, it turns out that Stratonovich integral fits better. See discussions in [MM22].

2 Itô Formula

Recall that in the example in Section 1.2, we have

$$Q_n = \sum_{i=1}^n (W(t_i) - W(t_{i-1}))^2, \quad \mathbf{E}[Q_n] = T \quad \text{and} \quad \mathbf{Var}[Q_n] \overset{n\to\infty}{\longrightarrow} 0.$$

Note that when $n \to \infty$, $\Delta_i W = \mathrm{d}W(t_{i-1})$. Then

$$\forall T,\ \int_0^T (\mathrm{d}W(t_{i-1}))^2 = T = \int_0^T \mathrm{d}t.$$

This suggests that $(\mathrm{d}W(t))^2 \approx \mathrm{d}t$ holds. By definition of $\{W(t)\}$, we know that $\mathrm{d}W(t) = W(t+\mathrm{d}t) - W(t) \sim \mathcal{N}(0, \mathrm{d}t)$. Let X and Y be two random variables that $X \sim \mathcal{N}(0, \mathrm{d}t)$ and $Y = X^2$. Then the formula $(\mathrm{d}W(t))^2 \approx \mathrm{d}t$ tells us that Y is well concentrated on $\mathrm{d}t$.

With this observation, we then (heuristically) deduce the chain rule under the definition of Itô integral.

2.1 *Classical Chain Rule of Differentiation*

In the classical context, for any two differentiable functions g and f, the chain rule of differentiation is

$$\frac{\mathrm{d}f(g(t))}{\mathrm{d}t} = f'(g(t)) \cdot g'(t).$$

This can be verified using the Taylor expansion:

$$\begin{aligned}
\mathrm{d}f(g(t)) &= f(g(t+\mathrm{d}t)) - f(g(t)) \\
&= f(g(t) + \mathrm{d}g(t)) - f(g(t)) \\
&= f'(g(t))\,\mathrm{d}g(t) + \frac{1}{2}f''(g(t))(\mathrm{d}g(t))^2 + \frac{1}{6}f'''(g(t))(\mathrm{d}g(t))^3 + o((\mathrm{d}g(t))^3).
\end{aligned}$$

Then it follows that

$$\frac{\mathrm{d}f(g(t))}{\mathrm{d}t} = f'(g(t)) \cdot g'(t) + o(\mathrm{d}t) = f'(g(t)) \cdot g'(t),$$

where $\mathrm{d}t$ tends to 0.

2.2 *The Chain Rule with Itô Integral*

Similarly, we have

$$\begin{aligned}
\mathrm{d}f(W(t)) &= f(W(t) + \mathrm{d}W(t)) - f(W(t)) \\
&= f'(W(t))\,\mathrm{d}W(t) + \frac{1}{2}f''(W(t))(\mathrm{d}W(t))^2 + \frac{1}{6}f'''(W(t))(\mathrm{d}W(t))^3 + o((\mathrm{d}W(t))^3).
\end{aligned}$$

This yields the Itô formula

$$\mathrm{d}f(W(t)) = f'(W(t))\,\mathrm{d}W(t) + \frac{1}{2}f''(W(t))\,\mathrm{d}t,$$

as $\mathrm{d}t \to 0$.

Consider a diffusion $\{X(t)\}$ that

$$\mathrm{d}X(t) = \mu(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t).$$

By the Itô formula, we have

$$\begin{aligned}
\mathrm{d}f(X(t)) &= f(X(t)+\mathrm{d}X(t)) - f(X(t)) \\
&= f'(X(t))\,\mathrm{d}X(t) + \frac{1}{2}f''(X(t))(\mathrm{d}X(t))^2 \\
&= f'(X(t))\,\mu(t,X(t))\,\mathrm{d}t + f'(X(t))\,\sigma(t,X(t))\,\mathrm{d}W(t) + \frac{1}{2}f''(X(t))(\sigma(t,X(t)))^2\,\mathrm{d}t
\end{aligned}$$

Then we see some examples of Itô formula.

Example 1 (Geometric Brownian Motion) *Recall the geometric Brownian motion $Y(t) = f(X(t))$ where f is the exponential function and $\{X(t)\}$ is a diffusion specified by $\mu(t,X(t)) \equiv 0$ and $\sigma(t,X(t)) \equiv 1$. Then it follows from the Itô formula that*

$$\mathrm{d}Y(t) = e^{X(t)}\left(\mathrm{d}W(t) + \frac{1}{2}\,\mathrm{d}t\right) = Y(t)\,\mathrm{d}W(t) + \frac{Y(t)}{2}\,\mathrm{d}t.$$

Example 2 (Ornstein-Uhlenbeck Process) *Let $\{X(t)\}$ be a Ornstein-Uhlenbeck process that $\mathrm{d}X(t) = -X(t)\,\mathrm{d}t + 2\,\mathrm{d}W(t)$. Let $f(t,x) = e^t \cdot x$. We adopt the following notations:*

$$\begin{aligned}
f_t(t_0,x_0) &= \frac{\partial f(t,x)}{\partial x}\bigg|_{t=t_0,x=x_0}, \\
f_{tx}(t_0,x_0) &= \frac{\partial}{\partial t}\frac{\partial}{\partial x}f(t,x)\bigg|_{t=t_0,x=x_0}.
\end{aligned}$$

and similarly define $f_x(t_0,x_0)$, $f_{xx}(t_0,x_0)$, $f_{tt}(t_0,x_0)$, and $f_{xt}(t_0,x_0)$. Then $f(t,X(t)) = e^t \cdot X(t)$ and using Taylor expansion, we have

$$\begin{aligned}
\mathrm{d}f(t,X(t)) &= f(t+\mathrm{d}t, X(t+\mathrm{d}t)) - f(t,X(t)) \\
&= f_t(t,X(t))\,\mathrm{d}t + f_x(t,X(t))\,\mathrm{d}X(t) \\
&\quad + \frac{1}{2}\left(f_{tt}(t,X(t))(\mathrm{d}t)^2 + 2f_{tx}(t,X(t))\,\mathrm{d}t\,\mathrm{d}X(t) + f_{xx}(t,X(t))(\mathrm{d}X(t))^2\right) \\
&= f_t(t,X(t))\,\mathrm{d}t + f_x(t,X(t))\,\mathrm{d}X(t) + \frac{1}{2}f_{xx}(t,X(t))(\mathrm{d}X(t))^2 + o(\mathrm{d}t)
\end{aligned}$$

Note that $f_{xx}(t,X(t)) = 0$. Thus

$$\begin{aligned}
\mathrm{d}f(t,X(t)) &= f_t(t,X(t))\,\mathrm{d}t + f_x(t,X(t))\,\mathrm{d}X(t) \\
&= e^t X(t)\,\mathrm{d}t + e^t\,\mathrm{d}X(t) \\
&= 2e^t\,\mathrm{d}W(t).
\end{aligned}$$

This yields that $\mathrm{d}(e^t X(t)) = 2e^t\,\mathrm{d}W(t)$ and sequentially

$$X(T) = e^{-T}\left(\int_0^T 2e^t\,\mathrm{d}W(t) + X(0)\right).$$

References

[Kle12] Fima C Klebaner. *Introduction to stochastic calculus with applications*. World Scientific Publishing Company, 2012. 3

[MM22] Riccardo Mannella and Peter VE McClintock. Itô versus stratonovich: 30 years later. *The Random and Fluctuating World: Celebrating Two Decades of Fluctuation and Noise Letters*, pages 9–18, 2022. 3