§ Stochastic differential equations

Let us begin with an ordinary differential equation

$$\frac{dX(t)}{dt} = a(X(t)), \quad t \in [0, \infty), \quad X(t_0) = X_0,$$

This model is useful in a plenty of applications. However, in reality models are intrinsically random, hence it is necessary to extend the deterministic ODE into a random one. Practioners and engineers would heuristically append a Gaussian-like noise on the RHS of the ODE above to get

right hand side

$$\frac{dX(t)}{dt} = a(X(t)) + \xi(t), \quad \text{LATEX \textbackslash xi.}$$

where de facto they usually demand

$$\boxed{\begin{array}{l} 1)\ \xi(t) \sim N(0, \sigma) \text{ is Gaussian } \forall t \in [0, \infty), \\ 2)\ \xi(t) \text{ and } \xi(s) \text{ are independent}, \ \forall t \neq s. \end{array}}$$

Indeed, when the heuristic random ODEs are simulated in discrete times, they indeed can introduce the randomness as wanted. (see, e.g., 'Lec2_ode_noisy.ipynb').

However, when the random ODE comes to the hand of a mathematician, the mathematician feels confused, because the equation does not make sense. suppose that such a process ξ exists, then, the path of ξ must be discontinuous everywhere. This violates the meaning of $\frac{dX(t)}{dt}$ which asks for an X to be differentiable everywhere, because a differentiable everywhere function cannot have discontinuous everywhere derivative. To solve this problem, we need to reinterpret the meaning of $\frac{d}{dt}$. ~~One to weaken~~

Remark 1. Reinterpratation of classical calculus is very commonly seen. ~~Foro~~ instance, the transport equation

$$\frac{\partial u(t,x)}{\partial t} = -\frac{\partial u(t,x)}{\partial x}, \quad t \in [0,\infty), \ x \in \mathbb{R}^d$$

$$u(0,x) = g(x).$$

Depending on the properties of the initial value g, it is possible that a differentiable solution u does not exist. To solve such problems, ~~ple~~ people invented weak derivatives to weaken $\frac{\partial}{\partial t}$ (see, e.g., Evans, 2010).

To avoid the differentiabity, we can rewrite the ODE into an ordinary integral equation

$$X(t) = X(t_0) + \int_0^t a(X(s))\, ds + \int_0^t h(s)\, ds$$

which does not explicity require X to be differentiable. However, this integral equation still does not make sense because the integral $\int_0^t h(s)\, ds$ is not well-defined. It is not well-defined because h is not Lebesgue-measurable. Even if we suppose that h is measurable, then we will ~~have~~ end up that $\int_0^t h(s)\, ds = 0$ amounts to zero almost surely for all t. Thus there is no randomness at all!

The reasons that h incurs these problems are that $h(t)$ has finite variance for every t, and that there are uncountably many of them are independent (see, Kallianpur, 1980, Example 1.2.5). Such a process h is a "faked" white noise process and is useless in practice. To introduce the "real" white noise process that we can use to randomise ODEs, the de facto choice is ~~that~~ the formal derivative of Brownian motion (also called Wiener process)

Definition 2. Brownian motion. A stochastic process W is called a standard Brownian motion / Wiener process if

1) for every $t, \Delta > 0$, the increment $W(t+\Delta) - W(t)$ is independent of $W(s)$ for all $s \le t$,

2) for every $t, \Delta > 0$, the increment $W(t+\Delta) - W(t) \sim N(0, \Delta)$ is Normal,

3) $W(0) = 0$ almost surely (a.s.)

4) $t \mapsto W(t)$ is continuoss a.s..

Essentially, a standard Brownian motion (BM) is a continuous-time Gaussian process with stationary and independent increments (imagine a limit of random walk). This property makes BM a suitable process for introducing noises in the ODE. We can write

$$\int_0^t d\overset{W(s)}{\cancel{B(s)}} \overset{\text{heuristically}}{\approx} \int_0^t w(s)\,ds,$$

where w - the formal derivative of W is a white noise process.

By using the BM, our stochastic ordinary integral equation (OIE) is

$$X(t) = X(t_0) + \int_0^t a(X(s))\,ds + \underbrace{\int_0^t dW(s)}_{:= W(t) \text{ by definition.}}$$

Essentially, instead of adding ξ to the ODE, we add W to the OIE. In what follows, we show that such stochastic integral ~~$\int_{t_0}^{t}$ is well-defined~~: Can be well-defined.

In reality, $\int_0^t dW(s)$ might not be enough to model noises; sometimes we also desire an integral of a function as

$$\int_0^t f(s)\,dW(s),$$

so the noises in the model are coupled by a function / process f. For simplicity, let us for now assume that the integrand f is deterministic. We show how to define $\int_0^t f(s)\,dW(s)$ in the celebrated Itô sense.

In the classical calculus, the integral $\int_0^t f(s)dW(s)$ really looks like a Riemann-Stieltjes integral. Hence, initially we would like to try to define the integral as

$W:[0,\infty)\times\Omega\to\mathbb{R}$

$(\Omega,\mathcal{F},\mathbb{P})$

$$I(f,\omega):=\int_0^t f(s)dW(s,\omega):=(RS)\int_0^t f(s)dW(s,\omega)$$

for all $\omega\in\Omega$ in the pathwise ~~way~~ fashion, viz., for every ω. we get a BM path, the we define the integral in the RS sense with respect to this path.

Example 3. Integrals, for example, $\int_0^t s\,dW(s)$, $\int_0^t \sin(s)dW(s)$ are well-defined in the pathwise RS ~~sens~~ definition. See, Kuo, 2006, Thm. 2.3.7.

However, the pathwise RS definition of stochastic integrals is ~~ambigeous~~ ambiguous, since the applicable integrands are very limited: $(RS)\int_0^t f(s)dW(s,\omega)$ makes sense only when f is continuous of bounded variation. BV The class of continuous bounded variation functions is quite narrow, hence, the integral is useful for a small family of integrands only (see, e.g., Example 3).

Moreover, if we want $(RS)\int_0^t f(s,\omega)dW(s,\omega)$ to make sense for all continuous integrands (not necessarily B.V.), then $t \mapsto W(t)$ must be of ~~B~~ bounded variation. However, this is not true, because $t \mapsto W(t)$ has infinite first order variation. we will detail this in ~~Lecture~~ the seminar Lecture 6.

So eventually, it is quite useless to define the stochastic integral in the pathwise RS sense. But please remark that this does not mean that we cannot define in the pathwise RS sense. Some textbooks have loose statements ~~that~~ claiming that we cannot, which is not true. of course we can, see, for instance, Example 3, it is just that it only applies to very limited integrands.

A much more useful definition of $\int_0^t f(s)\,dW(s)$ is that of Wiener integral. The Wiener integrals are special cases of Itô integrals, and ~~we~~ they ~~can~~ ~~a~~ are applicable for integrand $f \in L^2[0,T]$, where $L^2[0,T]$ stands for the space of square-integrable functions on $[0,T]$. Morover, the Wiener integral is independent of the sequence of step functions $\{f_n(t) := \sum_{i=1}^{n} c_i \mathbb{1}_{[t_{i-1},t_i)}(t)\}$ that $f_n \to f$, of course.

The Wiener integral is precisely a random variable defined by

$$\int_0^T f(s)\, dW(s) := \lim_{n\to\infty} \sum_{i=1}^{n} c_i \left(W(t_i) - W(t_{i-1}) \right)$$

in $L^2(\Omega)$

that converges in $L^2(\Omega)$ – a L^2 space of random variables with inner product $\langle X, Y\rangle := \mathbb{E}[XY]$. ~~No~~ Furthermore, for every $T>0$, the integral $\int_0^T f(s)\,dW(s)$ is a Normal random variable with mean and variance

$$\mathbb{E}\left[\int_0^T f(s)\,dW(s)\right] = 0,$$

$$\|I(f)\|^2_{L^2(\Omega)} = \mathbb{E}\left[\left(\int_0^T f(s)\,dW(s)\right)^2\right] = \int_0^T f(s)^2\,ds = \|f\|^2_{L^2[0,T]}.$$

Hence, the Wiener integral $I: L^2[0,T] \to L^2(\Omega)$ is an isometry.

~~(Recall Itô isometry)~~ (Cf. Itô isometry)

The Wiener integrals are defined for deterministic integrands. In reality, we also need to deal with random integrands, for instance $\int_0^t f(X(s))\,dW(s)$ for some adapted process X. To extend Wiener integrals for this case we need to use filtrations, stopping time and martingales which are not within the scope of this lecture note. Nevertheless, the two most used definitions are the Itô integral and Stratonovich integral. Throughout this Lecture we use the Itô integral, and for now ~~you~~ we can heuristically think of

$$\int_0^t f(X(s))\,dW(s) = \lim_{n\to\infty}\sum_{i=1}^{n} f(X(t_{i-1}))\big(W(t_i)-W(t_{i-1})\big),$$

where the choice of the point $X(t_{i-1})$ indicates that this is an Itô integral. In the Stratonovich sense, we ~~would~~ choose $X\left(\frac{t_i+t_{i-1}}{2}\right)$ which gives a different Limit. For details of the definition of these stochastic integrals, please see, for instance, Kuo, 2006.

By using the Itô integral, we can now ~~define~~ use the stochastic integral equation (SIE)

$$X(t) = X(0) + \int_0^t a(X(s))\,ds + \int_0^t b(X(s))\,dW(s),$$
$$X(0) = X_0 \sim P_{X_0}.$$

to be the shorthand of the stochastic differential equation (SDE)

$$dX(t) = \underbrace{a(X(t))}_{\text{drift}}\,dt + \underbrace{b(X(t))}_{\text{dispersion}}\,dW(t), \quad X(0) = X_0 \sim P_{X_0}.$$

In our context, the SIE and SDE are really the same object. We are essentially dealing with "integrals" instead of "differentials".

~~There are~~ The calculus that deals with the derivative of stochastic processes (e.g., $\frac{dX(t)}{dt}$) is called Malliavin calculus named after Paul Malliavin.

↳ Notes on solutions of SDEs.

We need a consensus on the meaning of solution existence and uniqueness of SDEs. In the context of ODEs, the existence can be such a function that is differentiable (up to certain ~~ord~~ degree) and satisfies the ODE. As for the ODE uniqueness, we can say that the solution is unique if any two solutions are pointwise equal.

In the SDE context, the notion of existence and uniqueness becomes trickier than that of ODEs. Let us recall the SDE

$$dX(t) = a(X(s))\,dt + b(X(s))\,d\boxed{W}(s),$$

The meaning of the notation $\boxed{W}$ above is actually can be unclear. Is this W a fixed/given BM or a placeholder for any BM? For instance, let W be a BM constructed by whatever means, then

- $W^{(2)}(t) := -W(t)$ is a BM,
- $W^{(3)}(t) := \int_0^t \mathrm{sgn}(W(s))\,ds$ is a BM
- $W^{(4)}(t) := W(t) - \int_0^t a(W(s))\,ds$ is a BM under another measure (see. Girsanov theorem, and let ~~b(x)~~ b be constant)
- $W^{(5)}(t)$ any continuous version of W
- .....

There are ~~a number~~ plenty ways to construct a BM. So back to the SDE. Does the W there specifically refer to a BM (e.g., $W^{(2)}, W^{(3)}$... above) or any BM? Suppose that we fix a specific BM, then the SDE solution X is such a process that is adapted to the generated sigma-algebra by the fixed BM and the initial condition and solves the SDE. In simple words, the solution $X(t)$ should be a function of $\{W(s): s \leq t\}$ and X_0. This definition of solution is usually referred as the strong solution.

on the other hand, if we interpret W in the SDE as a placeholder for any BM, we thus seek for a pair ~~of~~ solution (X, W) that satisfies the SDE. We call this weak solution. The weak solution is less restrictive because a strong solution is a weak solution but not vice versa, for instance, the Tanaka equation.

As for the uniqueness, since the solutions are stochastic processes, we have three notions. We say two solutions $X^{(1)}, X^{(2)}$ are unique pathwise if $\{X^{(1)}(t) = X^{(2)}(t),\ \forall t \in [0,\infty)\}$ a.s. We say two solutions are unique time-marginally, if $X^{(1)}(t) = X^{(2)}(t)$ a.s. for all $t \in [0,\infty)$. The pathwise and time-marginal uniquenesses coincide if the solutions are continuous. We say the solutions are unique in distribution, if $X^{(1)}$ and $X^{(2)}$ have the same finite-dimensional distribution.

We sometimes need more to have the SDE solutions satisfy certain properties, in particular, semimartingale, see, Karatzas and Shreve, 1991, pp. 285.

§ Itô's formula

Itô's formula is a fundamental tool in stochastic calculus ~~for semimartingales~~, as it is the chain rule for stochastic processes, in particular, semimartingales.

Let us recall the basic chain rule in the classical calculus. Let ϕ and X be two continuously differentiable functions, ~~denote $\phi, X \in C$.~~ The chain rule tells us that

we interchangably use notations $\frac{d\phi}{dX}(x)$ and $\frac{d\phi(x)}{dX}$, sorry!

$$\frac{d\phi(X(t))}{dt} = \frac{d\phi}{dX}\frac{dX(t)}{dt}.$$

The integral form of the chain rule reads

$$\phi(X(t)) - \phi(X(0)) = \int_0^t \frac{d\phi}{dX}\frac{dX}{dt}(s)\, ds = \int_0^t \frac{d\phi}{dX}(X(s))\, dX(s).$$

However, if the function X is a ~~stochastic process, the chain rule~~ the solution to an SDE, the chain rule generally does not hold, ~~because~~ The stochastic chain rule would have an additional term due to the finite non-zero quadratic variation. To see this, Let $\Delta_n := \{\overset{t_0=0}{t_0}, t_1, t_2, \ldots \overset{t_n=t}{t_n}\}$ be a partition of $[0, t]$. Then,

$$\phi(X(t)) - \phi(X(0)) = \sum_{i=1}^{n} \left(\phi(X(t_i)) - \phi(X(t_{i-1})) \right).$$

suppose that ϕ is twice-continuously differentiable, then by Taylor's theorem,

$$\phi(x) - \phi(z) = \phi'(z)(x-z) + \frac{1}{2}\phi''(c)(x-z)^2,$$

for some $c \in [x, z]$. Now, applying the result to, we get

$$\phi(X(t)) - \phi(X(0)) = \sum_{i=1}^{n} \phi'(X(t_{i-1}))(X(t_i) - X(t_{i-1})) + \frac{1}{2}\sum_{i=1}^{n} \phi''(c_i)(X(t_i) - X(t_{i-1}))^2,$$

where $c_i \in [X(t_{i-1}), X(t_i)]$.

Now let us take a limit of the partition, that is, $|\Delta_n| \to 0$

$\leftarrow := \max\{t_i - t_{i-1},\ i = 1, 2, \ldots n\}$

$$\phi(X(t)) - \phi(X(0)) = \lim_{|\Delta_n| \to 0} \sum_{i=1}^{n} \phi'(X(t_{i-1}))(X(t_i) - X(t_{i-1})) + \frac{1}{2}\lim_{|\Delta_n| \to 0} \sum_{i=1}^{n} \phi''(c_i)(X(t_i) - X(t_{i-1}))^2,$$

the first term ~~won~~ amounts to $\int_0^t \phi'(X(s))X'(s)\,ds = \int_0^t \phi'(X(s))\,dX(s)$,

and the second term amounts to zero, because the quadratic variation $\lim_{|\Delta_n| \to 0} \sum_{i=1}^{n} (X(t_i) - X(t_{i-1}))^2 = 0$.

proof sketch: Since X is continuously differentiable, there is a constant $\alpha = \sup_t |X'(t)|$, and $|X(t) - X(s)| \leq \alpha |t-s|$ (by mean value thm.). Hence,

$$\sum_{i=1}^{n} (X(t_i) - X(t_{i-1}))^2 \leq \alpha^2 \sum_{i=1}^{n} (t_i - t_{i-1})^2 \leq \alpha^2 \sum_{i=1}^{n} (t_i - t_{i-1})|\Delta_n|$$

$$\to 0 \text{ as } |\Delta_n| \to 0.$$

$\square$

This is why in classical calculus, we don't see the second term, because most used functions have zero quadratic variation.

However, it turns out that a lot of stochastic processes have non-zero quadratic variations. Take Brownian motion W for example, Itô, 1942 shows that $\sum_{i=1}^{n} (W(t_i) - W(t_{i-1}))^2 \xrightarrow{\text{in probability}} t$ and that

$$\cancel{\phi(B(t)) \to 0}$$

$$\phi(W(t)) - \phi(W(0)) = \int_0^t \phi'(W(s))\,dW(s) + \boxed{\frac{1}{2}\int_0^t \phi''(W(s))\,ds},$$

and the second term is really due to .

We have the following theorem for SDEs solutions.

Theorem 4 (Itô's formula). Let $x, t \mapsto \phi(x,t)$ be a function that is twice-differentiable and differentiable in its first and second arguments, respectively. Suppose that X is an Itô process solving the SDE

$$dX(t) = a(X(t))dt + b(X(t))\,dW(t), \quad X(0) = X_0,$$

$X \in \mathbb{R}$

then

$$\phi(X(t)) - \phi(X(0)) = \int_0^t \frac{\partial \phi}{\partial t}(X(s),s)\,ds$$
$$+ \int_0^t \frac{\partial \phi}{\partial x}(X(s),s)\,a(X(s)) + \frac{1}{2}\frac{\partial^2 \phi}{\partial x^2}(X(s),s)\,b(X(s))^2\,ds$$
$$+ \int_0^t \frac{\partial \phi}{\partial x}(X(s))\,b(X(s))\,dW(s).$$

The differential shorthand of the equation above is

$$d\phi(X(t)) = \left(\frac{\partial \phi}{\partial t}(X(t)) + \frac{\partial \phi}{\partial x}(X(t))\,a(X(t)) + \frac{1}{2}\frac{\partial^2 \phi}{\partial x^2}(X(t))\,b(X(t))^2\right)dt$$
$$+ \frac{\partial \phi}{\partial x}(X(t))\,b(X(t))\,dW(t),$$

that is, $\phi(X(t))$ is another Itô process solving the SDE above.

For multidimensional SDE, that is, $X \in \mathbb{R}^d$, $a: \mathbb{R}^d \to \mathbb{R}^d$, $b: \mathbb{R}^d \to \mathbb{R}^{d\times w}$, $W \in R^w$, $\phi: \mathbb{R}^d \times [0,\infty) \to \mathbb{R}$, the Itô's formula reads

$$\phi(X(t)) - \phi(X(0)) = \int_0^t \frac{\partial \phi}{\partial t}(X(s),s)\,ds$$

Later we may also write $\nabla_x \phi(X(s),s) \cdot a(X(s))$

$$+ \int_0^t \left(\nabla_x \phi(X(s),s)\right)^T a(X(s))$$
$$+ \frac{1}{2}\,\mathrm{tr}\left(\Gamma(X(s))\,H_x\phi(X(s),s)\right)ds$$
$$+ \int_0^t \left(\nabla_x \phi(X(s),s)\right)^T b(X(s))\,dW(s),$$

where $\nabla_x \phi$ is the gradient, $H_x\phi$ is the Hessian, and $\Gamma(x) := b(x)\,b(x)^T$.

Example 5. Let $\phi(x) = \sin(x)$ and

$$dX(t) = -X(t)dt + dW(t),$$

then.

$$\cancel{d\sin(x)} = d\sin(X(t)) = \left(-X(t)\cos(X(t)) - \frac{1}{2}\sin(X(t))\right)dt + \cos(X(t))\,dW(t)$$

Example 6. Let $\phi(x,t) = te^{x}$, and W be a Brownian motion,

then

$$e^{W(t)} - 0 = \int_0^t e^{W(s)}ds + \int_0^t \frac{s}{2}e^{W(s)}ds + \int_0^t se^{W(s)}dW(s)$$

§ Markov property

In spirit of Itô, the whole point of SDEs is really to construct continuos-time Markov processes. The Markov property is a powerful asset for cheap computations, the definition of which is informally given as follows.

Definition 7. (Markov process). A stochastic process X is said to be a Markov process on $[0, T]$, if

$$\mathbb{E}[\varphi(X(t+s)) \mid \underbrace{X(z) : z \le t}_{\text{this is informal, see, filtration.}}] = \mathbb{E}[\varphi(X(t+s)) \mid X(t)],$$

for every $t \in [0,T]$, $s \geq 0$, and bounded Borel measurable function φ (this includes, e.g., $\mathbb{1}_B$ for any B in the Borel sigma-algebra), or we can rewrite the expectations above in terms of distribution

$$\mathbb{P}_{X(t+s)}(\cdot \mid X(z) : z \le t) = \mathbb{P}_{X(t+s)}(\cdot \mid X(t))$$

The Markov property essentially says that to compute any statistical quantity of $X(t+s)$ based on all the past
Conditioned

information $\leq t$, we only need the information at the present time t.

Example 8. Consider a chain of random variables, defined by $X_k = f(X_{k-1}) + Q_k$, starting from X_0:

$$X_1 = f(X_0) + Q_1,$$
$$X_2 = f(X_1) + Q_2,$$
$$\vdots$$

It is obvious that $\mathbb{E}[X_k \mid X_{k-1}, X_{k-2}, \ldots X_0] = \mathbb{E}[X_k \mid X_{k-1}]$.

Solutions of SDEs are Markov processes (see, Kuo, 2006. ch.10). This is mainly because the driving term Brownian motion W is a Markov process. and the indepent increments of W. To see this, let us consider a simple example. Consider a Wiener integral $X(t) := \int_0^t f(z)\,dW(z)$. because of the independent increments of W, the random variable $\int_s^t f(z)\,dW(z)$ ~~is~~ for any $s<t$ is independent of $X(z)$ for all $z \leq s$. Hence,

$$
\begin{aligned}
&\mathbb{P}_{X(t)}(\cdot \mid X(z): z \le s) \\
&\left.\begin{aligned}
&= \mathbb{P}_{X(s)+\int_s^t f(z)\,dW(z)}(\cdot \mid X(z): z \le s) \\
&= \mathbb{P}_{X(s)+\int_s^t f(z)\,dW(z)}(\cdot \mid X(s))
\end{aligned}\right\} \text{why?} \\
&= \mathbb{P}_{X(t)}(\cdot \mid X(s))
\end{aligned}
$$

To make the Markov property easier, let us denote $p(x_k \mid x_j)$ the probability density function (PDF) of $X(t_k)$ conditioned on $X(t_j) = x_j$. The Markov property implies that

transition density / kernel

$$p(x_k \mid x_0, x_1, \dots x_{k-1}) = \boxed{p(x_k \mid x_{k-1})}.$$

This further implies that we can factorise ~~the~~ any finite-dimensional distribution of X in terms of the transition densities:

$$p(x_0, x_1, \dots x_k) = p(x_0) \prod_{i=1}^{k} p(x_i \mid x_{i-1}).$$

We also have

$$p(x_k \mid x_{k-1}) = \int p(x_k \mid x_j)\, p(x_j \mid x_{k-1})\, dx_j ,$$

for any $t_{k-1} < t_j < t_{k-1}$. This is known as the Chapman–Kolmogorov equation.

~~As an example~~ to see how the Markov property is useful in computation, consider an example as follows. Imagine that we want to ~~make samples~~ sampling $(X_1, X_2, \dots X_T) \sim p(x_1, x_2, \dots x_T)$ from the joint distribution. If T is large, we are sampling from a high-dimensional vector space $\mathbb{R}^T$. This is in general difficult. However, with the Markov property, we can sequentially do the sampling: First draw $X_1 \sim p(x_1)$, then $X_2 \mid X_1 \sim p(x_2 \mid x_1)$, ... and so on until $X_T \mid X_{T-1} \sim p(x_T \mid x_{T-1})$, which is much easier than directly drawn from the joint distribution.

As another example, suppose that $p(x_1, x_2, \dots x_T)$ is Normal, that is

$$p(x_1, x_2, \dots x_T) \propto \exp\Big(\underbrace{(x_{1:T} - \mu_{1:T})}_{\in\mathbb{R}^T,\ \in\mathbb{R}^T}{}^{T}\, \underbrace{C_{1:T}^{-1}}_{\in\mathbb{R}^{T\times T}} (x_{1:T} - \mu_{1:T}) \Big),$$

~~for some~~ In order to evaluate this joint PDF, we

have to tackle $C_{1:T}^{-1}$ $\in \mathbb{R}^{T\times T}$ which is expensive to compute.
However, with the Markov property, we will not have the matrix inversion. Please see. Exercise 1, Assignment 5.