Advanced computational methods-Lecture 1

This course:

Homework: 40%; Final exam: 60%. Alternative projects: This can be used as an alternative to the final exam. I will take the maximum as the 60% score.

No textbooks. I will upload lecture notes. I will setup a course website. Try to setup canvas.

Reference books

- Stochastic Numerics for Mathematical Physics. Milstein, Tretyakov.
- Numerical Solution of Stochastic Differential Equations. Loeden, Platen. Springer.
- Basics of Stochastic Analysis, by Timo Seppalainen
- Brownian motion and Stochastic calculus, by Karatzas, Shreve.
- Data Assimilation: A Mathematical Introduction, by K.J.H. Law, A.M. Stuart, K.C. Zygalakis

1 Motivation for SDEs

1.1 Dynamical systems with noise

In many problems, the full information is not accessible. For example, to model the motion of a pollen particle in water (the observation made by Robert Brown in 1828), the full force acting on the particle is unknown. Also, what governs the evolution of stock price is not fully clear. In all these problems, what we observe is oftens some "low dimensional dynamics" of the true system. In statistical physics, this is often called the open system, where there is energy exchange with the environment.

To model such systems, we consider an equation with noise.

$$\dot{X} = b(t, X) + \sigma(t, X)\eta(t)$$

where $\eta(t)$ is the noise. For the noise $\eta(t)$, we have some intuition:

- $\mathbb{E}\eta(t) = 0$
- $\eta(t_2)$ and $\eta(t_1)$ have the same distribution.
- $\mathbb{E}\eta(t)\eta(s) = \delta(t - s)$.

The third condition means that the the noise at different times are unrelated. We make the correlation to be Dirac delta so that the noise has nontrivial contribution for the system. The Fourier transform of the correlation function

$$\mathcal{F}(\delta(\cdot)) = 1.$$

This is independent of the frequency and we thus call $\eta(t)$ the (time-continuous) white noise.

Remark 1. *If the time is discrete, one can also define "white noise". For discrete white noise, the distribution may not be Gaussian. However, for the time-continuous white noise we condiser here, it must be a Gaussian process.*

Now, consider the integral of the white noise:

$$W(t) = \int_0^t \eta(s)\, ds.$$

This has the following properties

- $\mathbb{E}W = 0$
- $W(t)$ has stationary increments
- The increments are independent.
- $\mathbb{E}(W(t)W(s)) = s \wedge t.$

Moreover, if we expect W to have continuous paths, then $W(t)$ is in fact unique. This is called the Wiener process, or standard Brownian motion. Hence, the white noise is the derivative of Brownian motion.

Exercise: Using the above three properties to derive that

$$W(t) - W(s) \sim \mathcal{N}(t - s).$$

Remark 2. *In the theory of random fields, the white noise is generalized to a $L^2(\Omega)$-valued measure (i.e. a mapping from Borel sets to L^2 random variables). Then, $B_t = \eta([0, t]) \in L^2(\Omega)$.*

1.2 Stochastic differential equations in a formal way

With the understanding above, the dynamic system with noise can be written as

$$dX = b(t, X)dt + \sigma(t, X)dW,$$

or in integral form, we have

$$X(t) = X_0 + \int_0^t b(s, X(s))\, ds + \int_0^t \sigma(s, X(s))dW(s).$$

Such kind of equations are call stochastic differential equation (SDE) driven by Brownian motions. Of course, we have SDEs driven by other processes, which we will not touch in this course.

If σ is a constant, the integral of the last term gives $\sigma W(t)$, which is easy to understand. However, in general $\sigma(t, X)$ is random. We must understand how σ and dW are multiplied together. This will be answered by Itô integrals.

2 A glimpse of stochastic processes

The rigorous theory will be established on probability spaces using measure theory. Here, we look at these concepts briefly.

2.1 Preliminaries: notations and concepts

The *probability sapce/sample space* $(\Omega, \mathcal{F}, \mathbb{P})$ is used to caputre the randomness of stochastic phenomenon.

- Each $\omega \in \Omega$ is called a sample point and you can understand that it corresponds to one realization of the random phenomenon.
- $\mathcal{F}$ is the set (sigma algebra) of some subsets of Ω. These sets are called events.
- $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is the probability measure, which is a nonnegative measure with $\mathbb{P}(\Omega) = 1$.

A stochastic process $\{X_t\}$ is a family of random variables defined on $(\Omega, \mathcal{F}, P)$ indexed by t, taking values in a second measurable space, called the state space. In our course, the state space will usually be $\mathbb{R}^d$. Alternatively, we can also understand it as a $[0, \infty) \to S$ function valued random variable

$X(\cdot, \omega)$ on the probability space. Every such function is called a sample path (trajectory).

A process X is called *measurable* if for every $B \in \mathscr{B}$, $\{(t, \omega) : X_t(\omega) \in A\}$ is in $\mathscr{B}([0, \infty)) \otimes \mathcal{F}$. In this sense, X is a measurable function of two variables, i.e. $X = X(t, \omega)$.

We some times do not distinguish two processes. Here are different notions.

We say two stochastic processes X and Y are *indistinguishable* if their sample paths almost agree; more precisely:

$$\mathbb{P}[X_t = Y_t; \forall t < \infty] = 1.$$

This means that the set $\{\omega : X(\omega, t) = Y(\omega, t), \forall t\}$ has probability 1.

We say Y is a *modification* of X if for any $t \geq 0$,

$$\mathbb{P}[X_t = Y_t] = 1.$$

Since $[0, \infty)$ is uncountable, the second is strictly weaker than the first one.

Remark 3. *Of course, the index 't' sometimes are taken to be in $\mathbb{N}$, then one has the discrete time stochastic processes.*

To describe the information that an observer of a process knows, we use a family of σ-algebras. A filtration $\{\mathcal{F}_t : t \in \mathbb{R}_+\}$ is a collection of σ-algebras on the probability space such that

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F} \quad 0 \leq s < t < \infty.$$

For the convenience, we define

$$\mathcal{F}_\infty = \sigma(\cup_{t \geq 0} \mathcal{F}_t).$$

We say X is adapted to the filtration $\{\mathcal{F}_t\}$ if $X(t) \in \mathcal{F}_t$. (This means

$$\{\omega : X(t, \omega) \in B\} \in \mathcal{F}_t, \text{for all Borel set } B)$$

Often, we choose the minimal filtration for the process

$$\mathcal{F}_t = \sigma\{X_s : 0 \leq s \leq t\}$$

This represents all the information brought by X up to time t.

A stochastic process X is called *progressively measurable* with respect to the filtration $\{\mathcal{F}_t\}$ if for each $t \geq 0$ and $B \in \mathscr{B}$ (i.e. B is a Borel set), $\{(s, \omega) : 0 \leq s \leq t, X_s(\omega) \in A\}$ is in $\mathscr{B}([0, t] \otimes \mathcal{F}_t)$.

Proposition 1. *If X is measurable and adapted to a filtration, then it has a progressively measurable modification.*

Proposition 2. *If X is adapted to a filtration $\{\mathcal{F}_t\}$ and every sample path is right-continuous (or every sample path is left-continuous), then X is progressively measurable with respect to the filtration.*

The proof can be performed by observing that X is the pointwise limit of the sequence $X^{(n)}$ by right continuity (or left continuity) for any fixed $t > 0$.

$$X_s^{(n)} = X_{(k+1)t/2^n}(\omega), \ \forall s \in (\frac{kt}{2^n}, \frac{k+1}{2^n}t].$$

Exercise: fill in the details of the proof.

2.2 Stopping times

The motivation of is that we sometimes care about the occurance of a phenomenon. For example, the first time your amount of money exceeds 100,000 dollars. Intuitively, whether the event occurs prior to or at t should be part of the information known by the observer of the process at t. This then motivates the definition of *stopping times*.

Definition 1. *Let the measureable space $(\Omega, \mathcal{F})$ equipped with a filtration $\mathcal{F}_t$. A stopping time τ of the filtration is a random variable such that*

$$\{\omega : \tau \leq t\} \in \mathcal{F}_t, \ \forall t \geq 0.$$

An optional time T of the filtration is a random time T such that

$$\{\omega : T < t\} \in \mathcal{F}_t, \ \forall t \geq 0.$$

Exercise: If the filtration is right continuous, i.e. $\cap_{\epsilon>0}\mathcal{F}_{t+\epsilon} =: \mathcal{F}_{t+} = \mathcal{F}_t$, then any optional time is also a stopping time.

For more rigorous and complete discussion of filtrations, stopping times, read the book draft written by Timo, available on his website.

Definition 2. *Let τ be a stopping time. Then, the σ-field $\mathcal{F}_\tau$ is defined by*

$$\mathcal{F}_\tau = \{A : A \cap \{\tau \leq t\} \in \mathcal{F}_t\}$$

Proposition 3. *Consider two stopping times T, S.*

1. *If $A \in \mathcal{F}_S$, then $A \cap \{S \leq T\} \in \mathcal{F}_T$.*

2. *$\mathcal{F}_{S\wedge T} = \mathcal{F}_T \cap \mathcal{F}_S$. The events $\{S < T\}$, $\{S \le T\}$ and $\{S = T\}$ are all in $\mathcal{F}_T \cap \mathcal{F}_S$.*

The usefulness of stopping times is that we sometimes only care about the processes before a certain event happens. Techniquely, "stopped" processes will have better regularity.

Proposition 4. *Let X be a progressively measurable process for filtration $\{\mathcal{F}_t\}$, and τ is a stopping time. Then, the random variable $X_\tau := X_{\tau(\omega)}(\omega)$ is $\mathcal{F}_\tau$ measurable. Also, the stopped process $X_{\tau\wedge t}$ is progressively measurable.*

$X_{\tau\wedge t}$ is often a regularized process and it will be used to approximate X_τ as $t \to \infty$.

Example: consider a gambling game, wher the probability you win in each time is 1/2. You now use a strategy: Starting from betting 1 dollar, every time you lose, you double the bet in the next game until you win. Let τ be the stopping time which describes the first time you win. Then, $\tau < \infty$ a.s. and thus

$$X_\tau = 1, a.s.$$

This seems that you will always win. However, $X_{\tau\wedge t}$ is a process with

$$\mathbb{E}X_{\tau\wedge t} = 0, \ \forall t \in \mathbb{N}.$$

This means that at any finite time, you will not win. The expectation is zero! The tricky part is $\mathbb{E}\tau = +\infty$. Hence, unless the amount of the money you have is infinity and you play the game forever, otherwise, you cannot win. As a corollary, in this example, $X_{\tau\wedge t}$ converges almost surely to X_τ but not in L^1.

3 A biref introduction to Brownian motions

Definition 3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $\{\mathcal{F}_t\}$ is a filtration. An adapted real-valued process $\{B_t\}$ is called a one dimensional Brownian motion if*

1. *$t \mapsto B(t, \omega)$ is continuus almost surely.*

2. *$B_t - B_s$ is independent of $\mathcal{F}_s$ and normally distribution with mean zero, variance $t - s$.*

Moreover, if further $B_0 = 0$ almost surely, we call it the standard Brownian motion, or Wiener process, denoted also by $W(t)$.

If there is no filtration mentioned in the Brownian motion, then we use the filtration naturally associated with B:

$$\mathcal{F}_t = \sigma(B_s : 0 \leq s \leq t).$$

However, this filtration is not right continuous. Using some technique, one an enlarge it to be right continuous and the $\{B_s\}$ is also a Brownian motion for this new filtration.

The d-dimensional Brownian motion is similarly defined. We omit.

If you are interested in the basic constructions of Brownian motions and the Wiener measures, you may refer to section 2.2-2.4 in the book "Brownian motions and stochastic calculus".

Remark 4. *We say the Wiener process is the standard Brownian motion. In fact it is the specific one by one of the constructions here.*

One useful consequence from this construction is the Wiener measure defined on $C[0,\infty)$, denoted by $\mathbb{P}^0$. It is the measure induced by the standard Brownian motions. For example, let $A = \{f \in C[0,\infty) : f(t) \in E\}$, and $A_1 = \{f \in C[0,\infty) : f(0) = 0, f(t) \in E\}$. Then,

$$\mathbb{P}^0(A) = \mathbb{P}^0(A_1) = \frac{1}{\sqrt{2\pi t}} \int_E e^{-\frac{x^2}{2t}}\, dx.$$

The measure induced by Brownian motion by Brownian motions starting from x is denoted by $\mathbb{P}^x$. Hence,

$$\mathbb{P}^x(E) = \mathbb{P}^0(E - x).$$

Remark 5. *There is another Levy's theorem about Brownian motion that may be useful sometimes: If a martingale M has continuous paths and the quadratic variation is $[M,M] = t$, then M is the standard Brownian motion.*

3.1 Markov property and strong Markov property

The Brownian motion satisfies the Markov properties:

$$\mathbb{P}^\mu(X_{t+s} \in E|\mathcal{F}_s) = \mathbb{P}^\mu(X_{t+s} \in E|X_s)$$

Also,

$$\mathbb{P}^x[X_{t+s} \in E|X_s = y] = \mathbb{P}^y[X_t \in E],\ \mathbb{P}^x X_s^{-1},\ a.e.\ y.$$

As a corollary,

$$\mathbb{P}^x(\theta_t^{-1} A|\mathcal{F}_s) = \mathbb{P}^{X_t}(A),\ \mathbb{P}^x - a.s.,$$

where $\theta_s \omega(t) = \omega(t+s)$.

Moreover, it satisfies the strong Markov property. Strong Markov property basically says we can replace s by a stopping time S. We omit the details.

As a corollary, $B_{t+S} - B_S$ is a standard Brownian motion if S is a stopping time. (If S in general is a random time, the increment may not be a Brownian motion.)

3.2 The reflection principle

Let τ be a stopping time. Then, the probabilities that $B_{\tau+t}$ is above B_τ and below B_τ are equal. In fact, starting from any point, if you reflect the Brownian motion about this line, it is still a Brownian motion.

With this observation, we let $\tau_b = \inf\{t : B_t = b\}$ for $b > 0$. Then,

$$\mathbb{P}^0(\tau_b < t) = \mathbb{P}^0(\tau_b < t, B_t < b) + \mathbb{P}^0(B_t > b).$$

By the reflection principle (in fact, due to strong Markov property), it can be viewed as starting from τ_b and reach B_t. However, this is equal to the probability that it goes above B_{τ_b}. This means $\mathbb{P}^0(\tau_b < t, B_t < b) = \mathbb{P}^0(B_t > b)$.

Hence,

$$\mathbb{P}^0(\tau_b < t) = 2\mathbb{P}^0(B_t > b).$$

Exercise: derive the density of τ_b.

3.3 sample path properties

- $\frac{W_t}{t} \to 0$ by law of large numbers.
- The path is monotone in no interval.
- It is nowhere differentiable.
- $\limsup_{t\to 0} \frac{W(t)}{\sqrt{2t \log\log(1/t)}} = 1$ and $\limsup_{t\to +\infty} \frac{W(t)}{\sqrt{2t \log\log(t)}} = 1$. By symmetry, the lim inf will then be -1.
- The continuity is like $\sqrt{2h \log(1/h)}$. In other words,

$$\mathbb{P}\left[\limsup_{h\to 0^+} \frac{1}{\sqrt{2h \log(1/h)}} \sup_{|s-t|\le h, s,t\le 1} |W(t) - W(s)| = 1\right] = 1.$$

For other properties, like exit time, passage time, running maximum, read the book by Karatzas.