STOCHASTIC OPTIMIZATION IN CONTINUOUS TIME

FWU-RANQ CHANG

③ Stochastic calculus

A differential equation:

$$\dot{x} = \mu(t, x) \quad (3.1)$$

can be written as

$$dx = \mu(t, x)\, dt$$

and extended to be stochastic as

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t \quad (3.2)$$

$\uparrow$ Wiener process

What is a Wiener process?

See Xiao Lect 18

→ it's the limiting process of a random walk when you let the time interval go to zero

→ to me it seems more like the "3D analog" of WN

Note: a Wiener process is also a special case of Markov processes w/ normally distributed transition probability.

It is also known as Brownian motion, is continuous and nowhere differentiable.

"Big O" notation

$S = O(\sqrt{\Delta t})$ if $S \to 0$ at the rate $\sqrt{\Delta t}$, i.e.

$$\lim_{\Delta t \to 0} \frac{S}{\sqrt{\Delta t}} = k$$ for k constant.

"Little o" notation

$S = o(\Delta t)$ if $S \to 0$ faster than Δt

that is

$$\lim_{\Delta t \to 0} \frac{S}{\Delta t} = 0$$

For Wiener processes, even w/ Xiao we talked about the independent increment property, i.e.

$$W(t) - W(s) \perp W(s) - W(0)$$

In words: The step(s) the process takes btwn time t and s is/are independent from the ones between s and 0.

(Note that the RW fulfills this too.)

The probability density of a Wiener process is

$$f(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left\{ - \frac{(y-x)^2}{2t} \right\} \qquad (2.17)$$

But eq. (3.2) is not just "(3.1) + random shock"

So (3.2) does not represent the derivative of X_t wrt time, $\neq \frac{dX_t}{dt}$

What is the meaning then of (3.2)?

Note that a differential equation has an integral interpretation:

$\dot{x} = \mu(t,x)$ in (3.1) is equivalent to

$$X_t - X_0 = \int_0^t \mu(s,x)\, ds$$

i.e. "x evolves in time as $\mu(t,x)$" is equivalent to saying that the change in x between 0 and t is the sum of all the steps in $\mu(s,x)$ over that time horizon.

Similarly, $dX_t = \mu(t,X_t)dt + \sigma(t,X_t)\,dW_t$ in (3.2) is equivalent to

$$X_t - X_{t_0} = \int_{t_0}^t \mu(s,X_s)\,ds + \int_{t_0}^t \sigma(s,X_s)\,dW_s \quad (3.3)$$

which means that a sol. to (3.3) is a sol. to (3.2)!
(provided that $\int_{t_0}^{t} b(s, X_s)\, dW_s$ exists.)

1.) $\int_{t_0}^{t} b(s, X_s)\, dW_s$ is not a Riemann Integral.

Why? B/c a Riemann integral is one which you can write as a sum of Δ-s for the partitioned spaces between t_0 and t when the number of these partitions $\to \infty$.
But $\int_{t_0}^{T} b(t, X_t)\, dW_t$ is <u>not</u> independent of

the choice of intermediate points of a partition of $[t_0, T]$. The reason is that the subintegral

$\int_{t_0}^{T} W_T\, dW_t$ is also not Riemann integrable.

2.) Search for a class of functions $b(s, X_s)$

erm...

... all of this is leading up to the Ito integral which will somehow be the sol to (3.3).

The Ito Integral

p. 65

Probability space $(\Omega, \mathcal{F}, P)$.

<u>Def</u>. A family of σ-algebras $\{\mathcal{F}_t : t \in I\}$ is called a filtration i.e. an increasing family, if $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ whenever $s \leq t$.

In words: a filtration is a sequence of sets in which the most recent ones encompass their predecessors. E.g. info sets.

Point set Ω

The set of elementary events in prob. theory.

Power set 2^Ω

The set of all subsets of Ω.

Algebra / field

A class $\mathcal{F}$ of subsets of Ω (i.e. $\mathcal{F} \subset 2^\Omega$) if

(i) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

(ii) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$

(iii) $\Omega \in \mathcal{F} \Leftrightarrow \phi \in \mathcal{F}$

In words

(i) esik $\in \mathcal{F} \Rightarrow$ nem esik $\in \mathcal{F}$

(ii) esik $\in \mathcal{F}$ & fúj $\in \mathcal{F} \Rightarrow$ esik és fúj $\in \mathcal{F}$

(iii) ?

Def. A class $\mathcal{F}$ of subsets of Ω (- i.e. an algebra) is moreover a σ-algebra (or σ-field) if it also satisfies (i)-(iii) AND

(iv) if $A_i \in \mathcal{F}$ $i = 1, 2, \ldots$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

↳ i.e. this is an extension of (ii) to ∞ many possibilities. Mathematically we say that the union $\bigcup_{i=1}^{\infty} A_i$ is countable.

$\Rightarrow$ You can really think of an algebra (or a σ-algebra) as information sets: sets in probability theory w/ a particular structure

(i) if an event is in it, the opposite is also in it

(ii) if two dimensions are in it, then one or the other or both happening is also in it

(iv) (ii) holds for ∞ dimensions $\rightarrow$ σ-algebra.

A filtration additionally has this "encompassing in time" property.

Def. A set function $P : \mathcal{F} \to \mathbb{R}$ is a probability measure if P satisfies

(i) $0 \leq P(A) \leq 1 \quad \forall A \in \mathcal{F}$

(ii) $P(\emptyset) = 0$ and $P(\Omega) = 1$

↳ ez a "valami majd csak lesz" property

(iii) if $A_i \in \mathcal{F}$ and A_i are mutually disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

↳ (iii) is called countable additivity

The triplet $(\Omega, \mathcal{F}, P)$ are a probability space.

Borel sets

When $\Omega = \mathbb{R}$ or $\Omega = [0,1]$ and the σ-algebra is the one generated by the open sets in $\mathbb{R}$ or in $[0,1]$, then this σ-algebra is called the Borel field, $\mathcal{B}$.

An element in the Borel field is a Borel set.

When $\Omega = [0,1]$, the σ-algebra is $\mathcal{B}$, and $P(A)$ is the "length" (measure) of $A \in \mathcal{F}$, then P is the probability measure on $\mathcal{B}$ and is known as the Lebesgue measure on $[0,1]$.

It seems also as if being a Borel set meant that the set is $\in \mathbb{R}$ (or $[0,1]$) and is observable.

The Ito Integral . Cont.

Def. Ito integral of a step function $b(t,\omega)$ is

$$I(b)(\omega) = \int_{t_0}^{T} b(t,\omega)\, dW(t) = \sum_{i=0}^{n-1} b(t_i,\omega)\left[W(t_{i+1}) - W(t_i)\right]$$

i.e. it's the Riemann sum evaluated at the left endpoints. The Ito integral is a random variable.

Def. Some ass's + $\mu(t,\omega)$ is integrable

i.e. $\int_{t_0}^{T} |\mu(t,\omega)| \, ds < \infty$

so that

$$X_t = X_{t_0} + \int_{t_0}^{t} \mu(s,\omega) \, ds + \int_{t_0}^{t} \sigma(s,\omega) \, dW_s$$

then we say that $\mu(t,\omega)\,dt + \sigma(t,\omega)\,dW$ is the stochastic differential of the process and we denote it by dX_t. It is also called an Ito process.

Def. A process X_t is a martingale if

$$E[X_t | \mathcal{F}_s] = E[X_s | \mathcal{F}_s] = X_s \quad (3.12)$$

(p. 74)

I think for $t > s$.

Note: the Wiener process satisfies (3.12), and so it's a martingale.

$X_s \leq E[X_s | \mathcal{F}_s] \rightarrow$ submartingale

$X_s \geq E[X_s | \mathcal{F}_s] \rightarrow$ supermartingale

The Ito integral can be used to generate martingales.

Ito's Lemma . Autonomous case

(3.2) : $dX_t = \mu(t, X_t)\,dt + \sigma(t, X_t)\,dW_t$

is autonomous when the drift $\mu(t, X_t)$ and the instantaneous variance $\sigma(t, X_t)$ are independent of time:

$$dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t \qquad (3.13)$$

Ito's Lemma is essentially the stochastic version of the chain rule for differentiation. It answers the question of if X_t satisfies (3.13), and $Y_t = h(X_t)$, then what's dY_t?

In the deterministic case, if $\dot{x} = f(x)$, $y = h(x)$, then $\frac{dy}{dt} = \frac{d}{dt}\left[h(x)\right] = h'(x)\,\dot{x} = h'(x) f(x)$.

Ito's Lemma If $h(x)$ is twice differentiable, then

$$dY_t = h'(X_t)\left[\mu(X_t)\,dt + \sigma(X_t)\,dW_t\right] + \underline{\tfrac{1}{2} h''(X_t)\,\sigma^2(X_t)\,dt}$$

this 2nd order Taylor term shows up

Xiao in Lect 20 says (p. 5, Mac)

If $f \in C^2$, then for $f(W(t))$, where $W(t)$ is a BM, a continuous fct of a BM, it holds that

$$df(W(t)) = f'(W(t))\, dW(t) + \underline{\tfrac{1}{2} f''(W(t))\, dt}$$

Or, more generally, for $B(r) \sim BM(\omega^2)$

$$df(B(t)) = f'(B(t))\, d\, B(t) + \underline{\tfrac{\omega^2}{2} f''(B(t))\, dt}$$

.. and Xiao emphasizes that this 2nd order term in this "stochastic chain rule differentiation" is the reason why the stochastic integral (i.e. the stochastic differential equation) is <u>not</u> a Riemann integral, b/c for Riemann integrals, the 2nd order term would be negligible!

Note that we've made use of the following "multiplication table":

$\times$	dW	dt
dW	dt	0
dt	0	0

p. 79.

Def. geometric Brownian motion

$$dX_t = a X_t \, dt + b X_t \, dW_t \qquad (3.16)$$

And it has a closed-form sol:

$$X(t) = X(0) \cdot \exp\left\{\left(a - \frac{b^2}{2}\right)t\right\} \exp\left\{b W(t)\right\} \qquad (3.17)$$

Its discrete time counterpart is

$$X_{t+1} = (1 + a) X_t + b X_t \varepsilon_{t+1}$$

Prop. $E(X(t)) = X(0) e^{at}$ for $X_t \sim$ geom BM

$$Var(X(t)) = X^2(0) e^{2at}\left(e^{b^2 t} - 1\right)$$

Ex. Population dynamics is described by

$$\underset{\substack{\uparrow \\ \text{\# offspring} \\ \text{from person } i}}{X_i(t+h)} = \underset{\substack{\uparrow \\ \text{expected pop} \\ \text{growth rate}}}{nh} + \underset{\substack{\uparrow \\ \text{system-wide} \\ \text{shock}}}{z\,\eta(t,h)} + \underset{\substack{\uparrow \\ \text{idio shock}}}{v_i \varepsilon_i(t,h)}$$

thus, despite being discrete time, can be approximated using the geom. Brownian motion

$$\underset{\substack{\uparrow \\ \text{pop.}}}{dL} = n L \, dt + z L \, dW \qquad (3.19)$$

and the proposition allows us to characterize it.

Additive vs. multiplicative shocks

$$dL = nL\,dt + \sigma L\,dW \quad \text{(multiplicative)} \quad (3.19)$$

$$\Leftrightarrow \frac{dL}{L} = n\,dt + \sigma\,dW$$

$\Rightarrow$ then $L(t) > 0 \quad \forall t$

$$dL = nL\,dt + \sigma\,dW \quad \text{(additive)}$$

(change to pop. doesn't depend on level)

$\Rightarrow$ then $L(t) < 0$ w/ prob > 0, small.

To make life harder, the book now switches to the notation z_t^i for the Wiener process $W^i(t)$.

Multivariate Ito's Lemma

Let W_t^i be a Wiener process satisfying

$(dz_t)(dz_t') = \lambda\,dt$, $\quad \lambda = \text{corrcoeff}(z_t, z_t')$

If X_t follows (3.13): $dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dz_t$

and Y_t follows $dY_t = \nu(Y_t)\,dt + \theta(Y_t)\,dz_t'$, and

$Z_t = h(X_t, Y_t)$, $h \in C^2$, then

$$dZ_t = h_X(\mu\,dt + \sigma\,dz_t) + h_Y(\nu\,dt + \theta\,dz_t')$$
$$+ \Big[\frac{\sigma^2}{2} h_{XX}\,dt + \frac{\theta^2}{2} h_{YY}\,dt \quad + \lambda\sigma\theta\,h_{XY}\,dt \quad (3.22)$$

Non-autonomous Ito's Lemma

X_t follows (3.2):

$$dX_t = \mu(t, X_t)\,dt + \sigma(t, X_t)\,dZ_t$$

$H(t, X_t) \in C^{1,2}$. Then

$$dH(t, X_t) = \underbrace{H_t\,dt}_{} + H_x(\mu\,dt + \sigma\,dZ_t) + \frac{1}{2}\sigma^2 H_{xx}\,dt$$

new term compared to autonomous case.