Diffusion Models and SDEs

Lecture 2:

SDE Properties, Linear SDEs, Time Reversal and the h-transform

Teodora's Remark, from last Lecture

Last Lecture

The optimal predictor of X as a function of Y (Hilbert projection)

$$\arg\min_{f-\text{is measurable}} \quad \mathbb{E}\left(X - f(Y)\right)^2$$

Is given by the conditional expectation:

$$f^*(Y) = \mathbb{E}[X|Y]$$

Teodora's Remark, from last Lecture

Last Lecture

The optimal predictor of the future as a function of the past in a martingale:

$$\underset{f-\text{is measurable}}{\arg\min} \quad \mathbb{E}\left(X_{t+\delta} - f(X_t)\right)^2$$

Is given by past itself:

$$f^*(X_t) = \mathbb{E}[X_{t+\delta}|X_t] = X_t$$

SDE Properties

Quadratic Variation of Brownian Motion

$$\lim_{n\to\infty} \mathbb{E}\left(t - \sum_{i=1}^{n} (W_{t_{i+1}} - W_{t_i})^2\right)^2 = 0$$

SDE Properties

Quadratic Variation of Brownian Motion

$$\lim_{n\to\infty} \mathbb{E}\left(t - \sum_{i=1}^{n} (W_{t_{i+1}} - W_{t_i})^2\right)^2 = 0$$

	$\mathrm{d}W_t$	$\mathrm{d}t$
$\mathrm{d}W_t$	$\mathrm{d}t$	0
$\mathrm{d}t$	0	0

SDE Properties

Ito's Lemma

Given the SDE:

$$\mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t$$

Consider a function $f(t, x)$ doubly differentiable in space and admitting single derivatives in time. Then the process $Y_t = f(t, X_t)$ satisfies:

$$\mathrm{d}Y_t = \left(\partial_t f + \nabla f^\top \mu(X_t, t) + \frac{1}{2}\mathrm{tr}(\sigma(X_t, t)\nabla\nabla f \sigma(X_t, t))\right)\mathrm{d}t + \nabla f^\top \sigma(X_t, t)\mathrm{d}W_t$$

SDE Properties

Ito’s Lemma - Exercise : Geometric Brownian Motion

Let us solve the SDE:

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t$$

now consider the transformation $Y_t = \ln X_t$ what are ?

$$\partial_t f = ??, \quad \partial_x f = ?? \quad \partial_x^2 f = ??$$

SDE Properties

Ito’s Lemma - Exercise : Geometric Brownian Motion

Let us solve the SDE:

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t$$

now consider the transformation $Y_t = \ln X_t$ what are ?

$$\partial_t f = 0, \quad \partial_x f = 1/x \quad \partial_x^2 f = -1/x^2$$

$$\mathrm{d}Y_t = \left(\frac{\mu}{X_t} \cdot X_t - \frac{\sigma^2}{2X_t^2} \cdot X_t^2 \right) \mathrm{d}t - \frac{\sigma}{X_t} \cdot X_t \mathrm{d}W_t$$

SDE Properties

Ito's Lemma - Exercise : Geometric Brownian Motion

Let us solve the SDE:

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t$$

now consider the transformation $Y_t = \ln X_t$ what are ?

$$\partial_t f = 0, \quad \partial_x f = 1/x \quad \partial_x^2 f = -1/x^2$$

$$\mathrm{d}Y_t = \left(\mu - \frac{\sigma^2}{2}\right)\mathrm{d}t - \sigma \mathrm{d}W_t$$

SDE Properties

Ito's Lemma - Exercise : Geometric Brownian Motion

Now let us solve the SDE:

$$\mathrm{d}Y_t = \left(\mu - \frac{\sigma^2}{2}\right)\mathrm{d}t - \sigma \mathrm{d}W_t$$

$$Y_t = Y_0 + \left(\mu - \frac{\sigma^2}{2}\right)\int_0^t \mathrm{d}s - \sigma \int_0^t \mathrm{d}W_s = Y_0 + \left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t$$

Remember $Y_t = \ln X_t$ thus:

$$X_t = e^{Y_t} = X_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}$$

Fokker Plank Equation

How does the marginal density evolve (SDEs ⇔ Parabolic PDEs)

What is the probability density of the SDE solution at a given time ?

$$\mathrm{Law} X_t = p_t(x) = ???$$

There's a special PDE (think heat equation) whose solution yield the marginal density:

$$\partial_t p_t(x) = -\sum_{i=1}^{d} \partial_{x_i}[\mu_i(t, x_i) p_t(x)] + \sum_{i,j=1}^{d} \partial_{x_i, x_j}[\sigma\sigma_{ij}^\top(t, x) p_t(x)]$$

Fokker Plank Equation

How does the marginal density evolve (SDEs ⇔ Parabolic PDEs)

What is the probability density of the SDE solution at a given time ?

$$\mathrm{Law} X_t = p_t(x) = ???$$

There's a special PDE (think heat equation) whose solution yield the marginal density:

$$\partial_t p_t(x) = \mathcal{P}(p_t)$$

Infinitesimal Generator

Uniquely Characterises PDE and Adjoint to FPK Operator

Consider the following operator for a given SDE

$$\mathcal{A}_t[f(x)] = \lim_{t\to 0} \frac{\mathbb{E}[f(X_t)] - x}{t}$$

Can be shown to reduce to:

$$\begin{aligned}\mathcal{A}_t[f] &= \partial_t f + \mu \cdot \nabla f + \frac{1}{2}\sum_{ij}[\sigma\sigma^\top]_{ij}(x,t)\partial_{x_i,x_j} f \\ &= \partial_t f + \mathcal{P}^\dagger(f)\end{aligned}$$

Linear SDEs

OU - Process

Mean reverting process. Reverts you back to mu.

$$X_0 \sim \pi$$

$$\mathrm{d}X_t = \alpha(\mu - X_t)\mathrm{d}t + \sqrt{2\alpha}\mathrm{d}W_t$$

Linear SDEs

OU - Process

For simplicity focus on the 0-mean case.

$$X_0 \sim \pi$$

$$\mathrm{d}X_t = -\alpha X_t \mathrm{d}t + \sqrt{2\alpha}\mathrm{d}W_t$$

Linear SDEs

OU - Process

Can be solved analytically via Integrating factor + Ito’s Lemma (notice how X_t looks like the DDPM kernel):

$$X_t = X_0 e^{-\alpha t} + (1 - e^{-2\alpha t})^{1/2} W_1$$

$$X_t = X_0 e^{-\alpha t} + W_{1 - e^{-2\alpha t}}$$

Linear SDEs

OU - Process

Intuitively you can see how the limit behaves:

$$\lim_{t \to \infty} X_t \overset{??}{=} W_1 \sim \mathcal{N}(0, I)$$

This is a completely informal/heuristic treatment. Calling it a heuristic is kind, but you can see where it is going.

Linear SDEs

OU - Process

More formal arguments can be made:

$$||\mathrm{Law} X_t - \mathcal{N}(0, I)||_{\mathrm{TV}} \leq Ce^{-\alpha t}$$

Can be a bit tricky to show from scratch, typically involves working with the Fokker Plank Equation + Using an Eigen decomposition of its semi group. Alternatively, Martingale methods have also been used.

Non Linear SDEs - Simply Discretise

Euler Maruyama (EM) Discretisation

To solve SDEs of the form

$$\mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t$$

We simply discretize them via EM

$$X_0 \sim \pi,$$
$$\epsilon_{t_k} \sim \mathcal{N}(0, \gamma I)$$
$$X_{t_{k+1}} = X_{t_k} + \mu(X_{t_k}, t_k)\delta t + \sqrt{\delta t}\sigma(X_{t_k}, t_k)\epsilon_{t_k},$$

Can prove convergence in $\mathscr{L}^p(\mathbb{P})$. Can we design better integrators ?

Time Reversal - Chain Rule

A discrete time "heuristic" sketch

Consider trying to enforce the following equality,

$$p_{t|t+\delta}(x|y)p_{t+\delta}(y) \stackrel{(\delta)}{=} p_{t+\delta|t}(y|x)p_t(x)$$

When (this equality does not hold for delta > 0)

$$p_{t|t+\delta}(x|y) = \mathcal{N}(x|y + f^{+}(y)\delta, \delta\sigma^2)$$

$$p_{t+\delta|t}(y|x) = \mathcal{N}(x|y + f^{-}(y)\delta, \delta\sigma^2)$$

Time Reversal - Chain Rule

A discrete time "heuristic" sketch

Now take logs on both sides and re-arrange:

$$(f^+(x) + f^-(y))^\top (y - x) + \delta(||f^+(x)||^2 + |f^-(y)||^2) \overset{(\delta)}{=} \sigma^2(\ln p_{t+\delta}(y) - \ln p_t(x))$$

Now take limit in delta -> 0 (where the equality does hold):

$$(f^+(x) + f^-(y))^\top (y - x) = \sigma^2(\ln p_t(y) - \ln p_t(x))$$

Applying Taylors theorem, given some $\lim_{y \to x} h(y) = 0$ it follows that:

$$\begin{aligned}(f^+(x) + f^-(y))^\top (y - x) &= \ln p_t(y) - \ln p_t(x) \\ &= \sigma^2 \nabla \ln p_t(x)^\top (y - x) + h(y)^\top (y - x)\end{aligned}$$

Time Reversal - Chain Rule

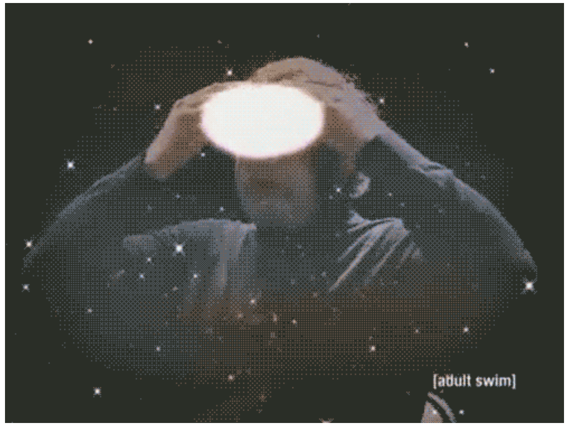
A discrete time “heuristic” sketch

Then, it follows that

$$f^{+}(y) + f^{-}(y) = \sigma^2 \nabla \ln p_t(y) + h(y)$$

Taking limit as x goes to why and invoking Taylors theorem once again (i.e. $\lim_{y \to x} h(y) = 0$) we get **Nelson’s duality**:

$$f^{+}(y) + f^{-}(y) = \nabla \ln p_t(y)$$



Time Reversal - Chain Rule

Formal Statement

Given the SDE

$$\mathrm{d}X_t = f^+(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$$

Then there exists an SDE:

$$\mathrm{d}Y_t = (\sigma_t^2 \nabla \ln p_{T-t}(Y_t) - f^+(Y_t, T-t))\mathrm{d}t + \sigma_t \mathrm{d}W_t$$

Such that:

$$Y_t \overset{(d)}{=} X_{T-t}$$

Doobs – Transform (Quick Version)

Formal Statement

Given the SDE

$$\mathrm{d}X_t = f(X_t, t)\mathrm{d}t + \sigma_t \mathrm{d}W_t$$

Then there exists an SDE:

$$\mathrm{d}Z_t = (f(Z_t, t) + \sigma^2 \nabla \ln p_{T|t}(z|Z_t))\mathrm{d}t + \sigma_t \mathrm{d}W_t$$

Such that Z_T goes through z:

$$Z_T \sim \delta_z$$

Appendix

Doobs – Transform (Quick Version)

Proof Sketch – Part I : Transition Density

First condition and apply Bayes Theorem

$$p_{t+\delta|t,T}(z_{t+\delta}|z_t, z_T = z) = \frac{p_{T|t,t+\delta}(z_T = z|z_t, z_{t+\delta})p_{t+\delta|t}(z_{t+\delta}|z_t)}{p_{T|t}(z_T = z|z_t)}$$

Now the Markov property

$$p_{t+\delta|t,T}(z_{t+\delta}|z_t, z_T = z) = \frac{p_{T|t+\delta}(z_T = z|z_{t+\delta})p_{t+\delta|t}(z_{t+\delta}|z_t)}{p_{T|t}(z_T = z|z_t)}$$

Now we need to find an SDE with this transition density.

Doobs – Transform (Quick Version)

Proof Sketch – Part 2 : Space Time Regular

The h-transform satisfies (since it satisfies backward Kolmogorov):

$$\mathcal{A}_t\left(p_{T|t+\delta}\left(z_T = z | z_{t+\delta}\right)\right) = 0$$

Doobs – Transform (Quick Version)

Proof Sketch – Part 3 : Finding the drift

Take time derivatives see what happens

$$\partial_t p_{t|s,T}(z_t|z_s, z_T = z) = \frac{1}{p_{T|s}(z_T = z|z_s)} \partial_t p_{T|t}(z_T = z|z_t) p_{t|s}(z_t|z_s)$$

$$= \frac{1}{p_{T|s}(z_T = z|z_s)} (p_{t|s}(z_t|z_s) \partial_t p_{T|t}(z_T = z|z_t) + p_{T|t}(z_T = z|z_t) \partial_t p_{t|s}(z_t|z_s))$$

$$= \frac{1}{h(z_s, s)} (-p_{t|s}(z_t|z_s) \mathcal{P}^\dagger h(z_t, t) + h(z_t, t) \mathcal{P} p_{t|s}(z_t|z_s))$$

$$= \frac{1}{h(z_s, s)} \left(\mathcal{P} h(z_t, t) p_{t|s}(z_t|z_s) + \nabla p_{t|s}(z_t|z_s) \cdot \nabla h(z_t, t) + p_{t|s}(z_t|z_s) \Delta h(z_t, t)\right)$$

$$= \frac{1}{h(z_s, s)} \left(\mathcal{P} h(z_t, t) p_{t|s}(z_t|z_s) + \nabla \cdot (h(z_t, t) p_{t|s}(z_t|z_s) \nabla \ln h(z_t, t))\right)$$