11/20 # Lecture 23: Girsanov's, Discretization

$\boxed{\text{Heuristic proof of Girsanov's:}}$

Given (A) $dx_t = b_t\, dt + \sqrt{2}\, dB_t$

(B) $dx_t = b'_t\, dt + \sqrt{2}\, dB_t$

Consider discrete-time approx, i.e.

$$\hat{x}_{(k+1)h} \leftarrow \hat{x}_{kh} + h \cdot b_{kh}(\hat{x}_{kh}) + \sqrt{2h}\, g_{kh}$$
$$\hat{x}_{(k+1)h} \leftarrow \hat{x}_{kh} + h \cdot b'_{kh}(\hat{x}_{kh}) + \sqrt{2h}\, g_{kh}$$

for $g_{kh} \sim N(0, I_d)$

Likelihood of observing trajectory

$$(\hat{x}_0, \hat{x}_h, \hat{x}_{2h}, \dots, \hat{x}_{Nh})$$

under (A) vs (B):

(A): $$\prod_{k=0}^{N-1} \exp\left(-\frac{1}{4h}\left\| \hat{x}_{(k+1)h} - \hat{x}_{kh} - h \cdot b_{kh}(\hat{x}_{kh})\right\|^2\right)$$

(B): $$\prod_{k=0}^{N-1} \exp\left(-\frac{1}{4h}\left\| \hat{x}_{(k+1)h} - \hat{x}_{kh} - h \cdot b'_{kh}(\hat{x}_{kh})\right\|^2\right)$$

$$\frac{A}{B} = \prod_{k=0}^{N-1} \exp\Big(-\frac{1}{4h}\Big[\|b_{kh}(\hat{x}_{kh})\|^2 \cdot h^2 - \|b'_{kh}(x_{kh})\|^2 \cdot h^2 - 2h\langle \underbrace{\hat{x}_{(k+1)h} - \hat{x}_{kh}}_{\substack{\text{under } B, \\ \text{this} = h \cdot b'_{kh}(\hat{x}_{kh}) + \sqrt{2h}\, g_{kh}}},\ b_{kh}(\hat{x}_{kh}) - b'_{kh}(\hat{x}_{kh})\rangle\Big]\Big)$$

$$= \prod_{k=0}^{N-1} \exp\Big(-\frac{1}{4h}\Big[-\|b_{kh}(\hat{x}_{kh}) - b'_{kh}(\hat{x}_{kh})\|^2 \cdot h^2 + h \cdot 2\sqrt{2}\langle \sqrt{h}\, g_{kh},\ b_{kh}(\hat{x}_{kh}) - b'_{kh}(\hat{x}_{kh})\rangle\Big]\Big)$$

$$= \exp\Big(-\frac{1}{4}\sum_{k=0}^{N-1} h\,\|b_{kh}(\hat{x}_{kh}) - b'_{kh}(\hat{x}_{kh})\|^2 + \frac{1}{\sqrt{2}}\sum_{k=0}^{N-1}\langle \underbrace{\sqrt{h}\, g_{kh}}_{\text{“}dB_{kh}\text{”}},\ b_{kh}(\hat{x}_{kh}) - b'_{kh}(\hat{x}_{kh})\rangle\Big)$$

$$\xrightarrow{(h\to 0)} \exp\Big(-\frac{1}{4}\int_0^T \|b_t - b'_t\|^2\, dt + \frac{1}{\sqrt{2}}\int_0^T \langle dB_t,\ b_t - b'_t\rangle\Big)$$

So

$$KL(\text{Ⓐ} \| \text{Ⓑ}) =$$

$$\mathbb{E}_{\text{Ⓐ}}\left[-\frac{1}{4}\int_0^T \|b_t - b'_t\|^2 dt + \frac{1}{\sqrt{2}}\int_0^T \langle \underbrace{dB_t}, b_t - b'_t \rangle\right] \quad (\star)$$

Brownian motion wrt. Ⓑ, so

$\mathbb{E}_{\text{Ⓐ}}[\cdot] \neq 0$!

Can also write dB_t in terms of Ⓐ by equating

$$b_t\, dt + \sqrt{2}\, d\hat{B}_t = b'_t\, dt + \sqrt{2}\, dB_t$$

$$\Rightarrow dB_t = \frac{1}{\sqrt{2}}(b_t - b'_t)\, dt + d\hat{B}_t.$$

Substituting into $(\star)$ and noting $\mathbb{E}_{\text{Ⓐ}}[d\hat{B}_t] = 0$,

$$KL(\text{Ⓐ} \| \text{Ⓑ}) = \frac{1}{4}\mathbb{E}_{\text{Ⓐ}}\int_0^T \|b_t - b'_t\|^2 dt \qquad \square$$

$\boxed{\text{Girsanov analysis for Langevin}}$

Suppose p is α-strongly-log-concave,
i.e. $\alpha \cdot Id \preceq -\nabla^2 \ln q \preceq L \cdot Id$

ALG: $d\hat{x}_t = -\nabla \ln q(\hat{x}_{kh})\,dt + \sqrt{2}\,dB_t$ for $t \in [kh, (k+1)h)$

TRUE: $dx_t = -\nabla \ln q(x_t)\,dt + \sqrt{2}\,dB_t$

$$KL(\text{TRUE} \| \text{ALG}) = \frac{1}{4} \mathbb{E}_{\text{TRUE}} \sum_{k=0}^{T/h - 1} \int_{kh}^{(k+1)h} \| \nabla \ln q(x_t) - \nabla \ln q(x_{kh}) \|^2 dt$$

$$\leq \frac{L^2}{4} \mathbb{E}_{\text{TRUE}} \sum_{k=0}^{T/h-1} \int_{kh}^{(k+1)h} \| x_{kh} - x_t \|^2 dt$$

$$\| x_t - x_{kh} \|^2 = \left\| \int_0^{t-kh} \nabla \ln q(x_{kh+s})\,ds + \sqrt{2}(B_t - B_{kh}) \right\|^2$$

$$\leq 2h \int_0^{t-kh} \| \nabla \ln q(x_{kh+s}) \|^2 ds + 4 \| B_t - B_{kh} \|^2$$

$$\lesssim L^2 h^2 \int_0^T \mathbb{E}_{\text{TRUE}} \| \nabla \ln q(x_t) \|^2 dt + L^2 h d T$$

Note

$$\mathbb{E}_{\text{TRUE}} \| \nabla \ln q(x_t) \|^2 \lesssim \mathbb{E}_q \| \nabla \ln q(x) \|^2 + \underbrace{L^2 W_2^2(\text{law}(x_t), q)}$$

$$\underset{\text{(Talagrand's } T_2\text{-inequality)}}{\leq} \frac{L^2}{\alpha} KL(\text{law}(x_t) \| q)$$

$$\underset{\text{(data processing)}}{\leq} \frac{L^2}{\alpha} KL(\text{law}(x_0) \| q)$$

Furthermore,

$$\mathop{\mathbb{E}}_{q} \|\nabla \ln q(x)\|^2 = \int \langle \nabla q, \nabla \ln q \rangle \, dx$$

$$\overset{(i.b.p.)}{=} -\int \Delta \ln q \cdot q \, dx$$

$$\overset{b/c\ -\nabla^2 \ln q \preceq L \cdot I_d}{\leq} Ld,$$

so $\forall t$,

$$\mathop{\mathbb{E}}_{TRUE} \|\nabla \ln q(x_t)\|^2 \lesssim Ld + \frac{L^2}{\alpha} KL(\mathrm{law}(x_0) \| q).$$

$$\Longrightarrow KL(TRUE \| ALG) \lesssim L^2 h^2 T \left(Ld + \frac{L^2}{\alpha} KL(\mathrm{law}(x_0) \| q)\right) + L^2 h d T$$

for sufficiently small ε, if we take $h = \Theta\left(\frac{\varepsilon^2}{L^2 d T}\right)$, then

$$KL(\mathrm{law}(x_T) \| \mathrm{law}(\hat{x}_T)) \leq KL(TRUE \| ALG) \lesssim \varepsilon^2$$

Also, $KL(\mathrm{law}(x_T) \| q) \leq KL(\mathrm{law}(x_0) \| q) \cdot \exp(-\Omega(\alpha T))$, because q satisfies log-Sobolev ineq. w/ constant $\Theta(\alpha)$, so if $T \gtrsim \frac{1}{\alpha} \lg(KL(\mathrm{law}(x_0) \| q) / \varepsilon)$, can ensure

$$KL(\mathrm{law}(x_T) \| q) \lesssim \varepsilon^2.$$

By Pinsker's inequality and triangle inequality for TV,

$$TV(\mathrm{law}(\hat{x}_T), q) \le TV(\mathrm{law}(\hat{x}_T), \mathrm{law}(x_T)) + TV(\mathrm{law}(x_T), q)$$

$$\lesssim \varepsilon. \qquad \square$$

Pros of this argument: discretization error bound only uses smoothness, i.e. the bound $-\nabla^2 \ln q \preceq L \cdot I_d$, not log-concavity

Cons: can only control discretization error over bounded time range

$\boxed{\text{Girsanov analysis for diffusion models}}$

ALG: $d\hat{x}_t = (\hat{x}_t + 2 s_{T-kh}(\hat{x}_{kh}))dt + \sqrt{2}\, dB_t$ for $t \in [kh, (k+1)h)$

TRUE: $dx_t = (x_t + 2\nabla \ln q_{T-t}(x_t))dt + \sqrt{2}\, dB_t$

$$KL(\mathrm{TRUE} \| \mathrm{ALG}) = \mathbb{E}_{\mathrm{TRUE}} \sum_{k=0}^{T/h-1} \int_{kh}^{(k+1)h} \|\nabla \ln q_{T-t}(x_t) - s_{T-kh}(x_{kh})\|^2 \, dt$$

$$\mathbb{E}\|\nabla \ln q_{T-t}(x_t) - s_{T-kh}(x_{kh})\|^2$$

$$\lesssim \mathbb{E}\|\nabla \ln q_{T-t}(x_t) - \nabla \ln q_{T-t}(x_{kh})\|^2 \qquad (1)$$

$$+$$

$$\mathbb{E}\|\nabla \ln q_{T-t}(x_{kh}) - \nabla \ln q_{T-kh}(x_{kh})\|^2 \qquad (2)$$

$$+$$

$$\mathbb{E}\| s_{T-kh}(x_{kh}) - \nabla \ln q_{T-kh}(x_{kh})\|^2 \qquad (3)$$

1 (discretization in space)

$$① \le L^2 \, \mathbb{E}\| x_t - x_{kh} \|^2$$

joint distribution agrees with that of forward process, i.e.

$$(x_t, x_{kh}) \underset{\text{law}}{=} \left(x_t,\ e^{-(t-kh)} x_t + N(0, 1-e^{-2(t-kh)} I_d)\right)$$

$$\le L^2 \left\{ \mathbb{E}\left\| \left(1 - e^{-(t-kh)}\right) x_t \right\|^2 + \underbrace{\left(1 - e^{-2(t-kh)}\right)}_{\Theta(h)} d \right\}$$

$$\lesssim L^2 h d + L^2 h^2 \, \mathbb{E}\|x_t\|^2$$

But $x_t \underset{\text{law}}{=} e^{-t} x + \sqrt{1-e^{-2t}}\, g$ for $x \sim q$, $g \sim N(0, I_d)$,

so $\mathbb{E}\|x_t\|^2 \le e^{-2t} m_2^2 + (1-e^{-2t}) d$

$$\le m_2^2 + d$$

$$\lesssim \boxed{L^2 h d + L^2 h^2 m_2^2}$$

3 (score error):

$$\boxed{③ \le \varepsilon_{sc}^2} \text{ by assumption}$$

② (discretization in time):

Sketch

Consider distribution $p \propto e^{-V}$ and $p' = p * N(0, \sigma^2 I_d)$ for $V \preceq L \cdot I_d$. Let's see how to control

$$\mathbb{E}_{p'} \left\| \nabla \ln p - \nabla \ln p' \right\|^2$$

using that $(a * b)' = a' * b$

Note

$$\nabla \ln p'(x) = \frac{\int p(y) \exp\left(-\frac{1}{2\sigma^2}\|y-x\|^2\right) \cdot \frac{\nabla p(y)}{p(y)}\, dy}{\int p(y) \exp\left(-\frac{1}{2\sigma^2}\|y-x\|^2\right) dy}$$

$$= -\mathbb{E}_{P_{x,\sigma^2}}\left[\nabla V(y)\right]$$

for $P_{x,\sigma^2} \triangleq \text{law}(y \mid y + \sigma g = x)$ for $y \sim p$

$$\mathbb{E}_{x\sim p'} \left\| \nabla \ln p(x) - \nabla \ln p'(x) \right\|^2 = \mathbb{E}_{x\sim p'} \left\| \mathbb{E}_{y\sim P_{x,\sigma^2}} \left[\nabla V(y) - \nabla V(x)\right] \right\|^2$$

$$\leq L^2 \cdot \mathbb{E}_{x\sim p'} \mathbb{E}_{y\sim P_{x,\sigma^2}} \|y - x\|^2$$

But $\text{law}(x, y) = \text{law}(\hat{y} + \sigma g, \hat{y})$ for $\hat{y} \sim p$,

$$\leq L^2 \sigma^2 d$$