Sampling and Diffusion

Wednesday, January 15, 2025 7:48 AM

plan

(LD) $\quad dX_t = -\nabla f(X_t)\,dt + \sqrt{2}\,dW_t$

We saw last time that

$$p(X_t) \longrightarrow e^{-f}$$

and converges in W_2 distance for strongly convex f.

The proof uses the Fokker-Planck equation.

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

has $\quad \dfrac{dp_t}{dt} = -\nabla\cdot(p\mu) + \dfrac{1}{2}\nabla\cdot\left(\nabla\cdot\left(\sigma\sigma^T\mu\right)\right)$

To prove this we need to understand how functions of X_t change.

For a variable x we have $\dfrac{df(x)}{dt} = f'(x)\dfrac{dx}{dt}$

$f(x+hv) = f(x) + h\langle\nabla f(x), v\rangle + \underline{O(h^2)} \qquad (dx)^2 \longrightarrow 0.$

$dt = h \to 0 \qquad$ what about stochastic X?

Lemma (Itô) $dX_t = \mu(X_t)\,dt + \sigma(X_t)\,dW_t$

$\quad f'(X_t)\,dX_t + \frac{1}{2} f''(X_t)\,\sigma(X_t)^2\,dt$

Lemma (Ito) $dX_t = \mu(X_t)\,dt + \ldots$

(1 dim) $$df(X_t) = f'(X_t)\,dX_t + \frac{1}{2} f''(X_t)\,\sigma(X_t)^2\,dt$$

$X_t, \mu_t \in \mathbb{R}^n$

$\sigma_t \in \mathbb{R}^{n\times m}$

$W_t \in \mathbb{R}^m$

$$df(X_t) = \langle \nabla f(X_t), dX_t\rangle + \frac{1}{2} \nabla^2 f(X_t) \cdot \sigma(X_t)\sigma(X_t)^T\,dt$$

$W_t \sim N(0,t)$

$t' > t.\ W_{t'} - W_t \sim N(0, t'-t)$

informally $(dW_t)^2 = (W_{dt})^2 = dt$

So need to keep second order term in Taylor expansion

$$df(X_t) = \langle \nabla f(X_t), dX_t\rangle + \frac{1}{2} dX_t^T\, \nabla^2 f(X_t)\, dX_t$$

$$= \langle \nabla f(X_t), dX_t\rangle + \frac{1}{2}\mu(X_t)^T \nabla^2 f(X_t)\mu(X_t)(dt)^2 \longrightarrow 0$$

$$+ \frac{2}{2}\mu(X_t)^T \nabla^2 f(X_t)\,\sigma(X_t)\,dt\,dW_t \longrightarrow 0$$

$$+ \frac{1}{2}\sigma(X_t)^T\, \nabla^2 f(X_t)\,\sigma(X_t)\,(dW_t)^2$$

$$= \langle \nabla f(X_t), dX_t\rangle + \frac{1}{2}\langle \nabla^2 f(X_t), \sigma(X_t)\sigma(X_t)^T\rangle\,dt.$$

E.g., $dX_t = dW_t$

$f(x) = \|x\|^2$

$\nabla f(x) = 2x$

$\nabla^2 f(x) = 2I$

$$d\|X_t\|^2 = \langle 2X, dW_t\rangle + \frac{1}{2}\langle 2I, I\rangle\,dt$$

$$= 2X^T dW_t + n\,dt$$

$$\mathbb{E}(\|X_t\|^2) = \mathbb{E}\left(\int_0^t 2X^T dW_t\right) + nt + \|X_0\|^2$$

This is a chain rule for stochastic variables.

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

$$df(X_t) = \sum_i \frac{\partial f}{\partial x_i}(X_t)\,dX_t^i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 f(X_t)}{\partial x_i \partial x_j}\,d[X^i, X^j]_t$$

$$dX_t^i = \mu(X_t)_i\,dt + \sigma(X_t)_i\,dW$$

$$d[X^i, X^j]_t = \left[\sigma(X_t)\sigma(X_t)^T\right]_{ij}\,dt$$

Before we prove F-P, let's review a basic technique in Calculus.

Integration by parts.

$$\int u\,dv = uv - \int v\,du$$

$$\nabla\cdot = \sum_i \frac{\partial}{\partial x_i} \qquad \text{divergence}$$

More generally, $\phi : \mathbb{R}^n \to \mathbb{R}$ (function), $u : \mathbb{R}^n \to \mathbb{R}^n$ (vector field)

$$\nabla\cdot(\phi u) = \nabla\phi\cdot u + \phi(\nabla\cdot u)$$

$$\int_\Omega \nabla\phi \cdot u = \int_\Omega \nabla\cdot(\phi u) - \int_\Omega \phi(\nabla\cdot u)$$

"divergence theorem":

$$= \int_{\partial\Omega} \phi u \cdot \vec{n} - \int_\Omega \phi(\nabla\cdot u)$$

<u>Pf.</u> (F-P) $X_0 \sim p$ $X_t \sim p_t$.

ϕ: smooth function.

$$\mathbb{E}_{x\sim p_t}(\phi(x)) = \mathbb{E}_{x\sim p}(\phi(X_t))$$

$$\int_\Omega \phi(x)\, p_t(x)\, dx = \int_\Omega \phi(X_t)\, p(x)\, dx$$

Time differentia

$$\int \phi(x)\, dp_t(x)\, dx = \int d\phi(X_t)\, p(x)\, dx \qquad \text{By Itô}$$

$$= \int \langle \nabla\phi(X_t), dX_t\rangle\, p(x)\, dx + \frac{1}{2}\int \langle \nabla^2\phi(x_t), \sigma(X_t)\sigma(X_t)^T\rangle\, p(x)\, dt\, dx$$

$$= \int \langle \nabla\phi(X_t), \mu(X_t)\rangle\, p(x)\, dt\, dx + \int \langle \nabla\phi(X_t), \sigma(X_t)\rangle\, p(x)\, dW_t\, dx$$

$$+ \frac{1}{2}\int \langle \nabla^2\phi(X_t), \sigma(X_t)\sigma(X_t)^T\rangle\, p(x)\, dt\, dx$$

First term:

$$\mathop{\mathbb{E}}_{x\sim p}\left(\langle \nabla\phi(X_t), \mu(X_t)\rangle\right) = \mathop{\mathbb{E}}_{x\sim p_t}\left(\langle \nabla\phi(x), \mu(x)\rangle\right)$$

$$= \int_{\Omega} \langle \nabla\phi(x), \mu(x)\rangle\, p_t(x)\, dx$$

$$= \underbrace{\int_{\partial\Omega} \phi(x)\mu(x)\cdot p_t(x)\vec{n}}_{=0} - \int \phi(x)\left(\nabla\cdot \mu(x)p_t(x)\right) dx$$

$p_t(x) \to 0$.

Second term: take expectation over process $\mathbb{E}(dW_t) = 0$

So $= 0$.

Third term $\int \langle \nabla^2\phi(x), \sigma(x)\sigma(x)^T\rangle\, p_t(x)\, dx$

I-by-P: $= -\int \langle \nabla\phi(x), \nabla\cdot\left(\sigma(x)\sigma(x)^T p_t(x)\right)\rangle\, dx$

I-by-P: $= \int \phi(x)\; \nabla\cdot\left(\nabla\cdot\left(\sigma(x)\sigma(x)^T p_t(x)\right)\right)\, dx$

So

$$\int \phi(x)\left[\frac{dp_t(x)}{dt} + \nabla\cdot\left(p_t(x)\mu(x)\right) - \frac{1}{2}\nabla\cdot\left(\nabla\cdot\left(\sigma(x)\sigma(x)^T p_t(x)\right)\right)\right] dx = 0$$

For all smooth ϕ. So integrand is 0. Done!

How about KL-divergence? $\nu = e^{-f}$

$$d_{KL}(\rho, \nu) = H_\nu(\rho) = \int \rho \log \frac{\rho}{\nu}$$

$$\frac{d}{dt} H_\nu(\rho) = \int \frac{d\rho}{dt} \log \frac{\rho}{\nu} + \cancel{\rho} \cdot \frac{\nu}{\cancel{\rho}} \cdot \frac{1}{\cancel{\nu}} \cdot \frac{d\rho}{dt}$$

$$= \int \frac{d\rho}{dt} \left(\log \frac{\rho}{\nu} + 1 \right) = \int \frac{d\rho}{dt} \left(\log \frac{\rho}{\nu} \right) + \underbrace{\frac{d}{dt} \int \rho}_{=0}$$

$$\boxed{\frac{d\rho}{dt} = \nabla \cdot \left(\rho \nabla \log \frac{\rho}{\nu} \right)}$$ By F-P.

$$= \int \nabla \cdot \left(\rho \nabla \log \frac{\rho}{\nu} \right) \log \frac{\rho}{\nu}$$

$$= - \int \left\langle \rho \nabla \log \frac{\rho}{\nu}, \nabla \log \frac{\rho}{\nu} \right\rangle$$

$$= - \int \rho \left\| \nabla \log \frac{\rho}{\nu} \right\|^2 = - \mathbb{E}_\rho \left(\left\| \nabla \log \frac{\rho}{\nu} \right\|^2 \right)$$

relative Fisher information $J_\nu(\rho)$.

Log-Sobolev Inequality (LSI)

$$2 H_\nu(\rho) \le C_{LSI} \cdot J_\nu(\rho)$$ holds for all ρ.

This is LSI for ν with constant C_{LSI}.

This is LSI for ν with constant C_{LSI}

As a result, for target ν with C_{LSI},

$$\frac{d}{dt} H_\nu(\rho_t) \le -J_\nu(\rho_t) \le -\frac{2}{C_{LSI}} H_\nu(\rho_t)$$

$$\Rightarrow \boxed{H_\nu(\rho_t) \le e^{-\frac{2t}{C_{LSI}}} \cdot H_\nu(\rho_0)},$$

<u>Thm</u>. For μ-strongly convex f, $\nu = e^{-f}$, $C_{LSI} \le \frac{1}{\mu}$.

$$\therefore H_\nu(\rho_t) \le e^{-2\mu t} H_\nu(\rho_0).$$

<u>LSI</u>: $\quad 2 H_\nu(\rho) \le C_{LSI} J_\nu(\rho)$

$\forall \rho$: $\quad \int \rho \log \frac{\rho}{\nu} \le \frac{C_{LSI}}{2} \int \rho \|\nabla \log \frac{\rho}{\nu}\|^2$

Equivalently: $\forall$ smooth g

$$\mathrm{Ent}_\nu(g^2) \le C_{LSI} \mathbb{E}_\nu(\|\nabla g\|^2)$$

$$\mathbb{E}_\nu(g^2 \log g^2) - \mathbb{E}_\nu(g^2) \log \mathbb{E}(g^2) \le C_{LSI} \mathbb{E}_\nu(\|\nabla g\|^2)$$

[...] the equivalence.

To see the equivalence,

$$g = \sqrt{\frac{\rho}{\nu}} \qquad g^2 = \frac{\rho}{\nu} \qquad \mathbb{E}_\nu(g^2) = 1.$$

$$\int \frac{\rho}{\cancel{\nu}} \log \frac{\rho}{\nu} \cdot \cancel{\nu} - 0 \le C_{LSI} \int$$

$$\nabla g^2 - 2g\nabla g = \frac{\nabla \rho}{\nu}$$

$$\|\nabla g\|^2 = \frac{\nu}{2\rho} \cdot \left\| \frac{\nabla \rho}{\nu} \right\|^2$$

$$= \frac{1}{2} \frac{\nu^2}{\rho^2} \left\| \frac{\nabla \rho}{\nu} \right\|^2 \frac{\rho}{\nu} = \frac{1}{2} \left\| \nabla \log \frac{\rho}{\nu} \right\|^2 \frac{\rho}{\nu}$$

So $$\int \rho \log \frac{\rho}{\nu} \le \frac{C_{LSI}}{2} \int \rho \left\| \nabla \log \frac{\rho}{\nu} \right\|^2$$

For the other direction, set $\rho = \dfrac{g^2}{\mathbb{E}_\nu g^2}$

$$\int_\nu \frac{g^2}{\mathbb{E} g^2} \log \frac{g^2}{\mathbb{E}(g^2)\nu} \le \frac{C_{LSI}}{2} \int_\nu \frac{g^2}{\mathbb{E} g^2} \left\| \nabla \log \frac{g^2}{\mathbb{E} g^2} \right\|^2$$

$$\int_\nu g^2 \log g^2 - \int_\nu g^2 \log \int_\nu g^2 \le \frac{C_{LSI}}{2} \int_\nu g^2 \cdot \frac{2}{g^2} \|\nabla g\|^2$$

$$\mathrm{Ent}(g^2) \le C_{LSI}\, \mathbb{E}_\nu(\|\nabla g\|^2)$$

$$\mathbb{E}\mathrm{nt}_{\nu}(g^2) \le C_{LSI}\, \mathbb{E}_{\nu}\left(\|\nabla g\|^2\right)$$