Lecture 7: Continuous-time samplers

- ☐ Video + discussion
- ☐ Langevin dynamics (MALA)
- ☐ Underdamped Langevin dynamics (ULA)
- ☐ Hybrid Monte Carlo (HMC)

Observations from video

1. Gibbs sampler makes big updates (50% of the variables at a time), moves fastest.
2. Methods with momentum (HMC/NUTS) do better than those without (MALA)
3. Choosing a step size in HMC is tricky, NUTS is HMC with adaptive step sizing.

MALA To sample $\pi(x)$ density, use

$$dX = \nabla \log \pi(x)\, dt + \sqrt{2}\, dW$$

- By definition of SDE as limit of Euler discretization:

$$X_{t+\delta} \approx X_t + \delta \cdot \nabla \log \pi(X_t) + (W_{t+\Delta t} - W_t)$$

- Question: how does the density evolve?

Let ρ_t denote the density of X_t.

Then, for bounded, smooth functions f,

$$\lim_{t\to 0} \frac{\int \rho_t(x) f(x)\,dx - \int \rho_0(x) f(x)\,dx}{t}$$

$$= \mathbb{E}\left[\lim_{t\to 0} \frac{f(X_t)-f(X_0)}{t} \,\middle|\, X_0 \sim \rho_0 \right]$$

$$= \mathbb{E}\left[\lim_{t\to 0} \frac{1}{t}\left[\nabla f(X_0)^T (X_t - X_0) + \frac{1}{2}(X_t - X_0)^T \nabla^2 f(X_0)(X_t - X_0)\right] \,\middle|\, X_0 \sim \rho_0 \right]$$

$$= \lim_{t\to 0} \frac{1}{t}\mathbb{E}\left[\nabla f(X_0)^T \left(t\cdot\nabla \log \pi(X_0) + \sqrt{2}\, W_t\right) \right.$$

$$\left. + \frac{1}{2}\left(t\cdot\nabla\log\pi(X_0) + \sqrt{2}W_t\right)\nabla^2 f(X_0)\left(t\cdot\nabla\log\pi(X_0) + \sqrt{2}W_t\right) \,\middle|\, X_0 \sim \rho_0 \right]$$

$$= \mathbb{E}\left[\nabla f(X_0)^T \nabla \log \pi(X_0) + \mathrm{tr}\left(\nabla^2 f(X_0)\right) \,\middle|\, X_0 \sim \rho_0 \right]$$

$$= \int \left[\mathrm{tr}\left(\nabla^2 \rho_0(x)\right) - \nabla\cdot\left(\rho_0(x)\nabla\log\pi(x)\right)\right] f(x)\,dx$$

whence

$$\frac{d}{dt}\rho(x) = \mathrm{tr}\left(\nabla^2\rho(x)\right) - \nabla\cdot\left(\rho(x)\nabla\log\pi(x)\right)$$

$$= \nabla\cdot\left(\rho(x)\nabla\log\left(\frac{\rho(x)}{\pi(x)}\right)\right).$$

$\Rightarrow$ There is a unique invariant measure satisfying $\frac{d}{dt}\rho(x) = 0$. What is it?

$$\rho(x) = \pi(x) \; !$$

$\Rightarrow$ $dX = \nabla \log \pi(x)\, dt + \sqrt{2}\, dW$ converges to invariant measure $\pi(x)$!

Ex Want to sample $\mathcal{N}(0, \Sigma)$.

$$\Rightarrow \quad \pi(x) = \frac{1}{|2\pi\Sigma|^{1/2}} \exp\left(-\tfrac{1}{2} x^T \Sigma^{-1} x\right)$$

$\Rightarrow$ Continuous-time sampler

$$dX = -\Sigma^{-1} X\, dt + \sqrt{2}\, dW$$

$\Rightarrow$ Discrete-time sampler

$$X_{t+\Delta} = \left(I - \delta \cdot \Sigma^{-1}\right) X_t + \sqrt{2 \cdot \delta}\; \xi_t$$

where $\xi_t \sim \mathcal{N}(0, I)$.

Q: How fast does this converge?

A: We can multiply by an eigenvector of Σ^{-1} with eigenvalue λ_i^{-1} and look at the resulting (one-dimensional) dynamics:

$$y_{t+\Delta} = \left(1 - \frac{\delta}{\lambda_i}\right) y_t + \sqrt{2\delta}\, z_t, \quad z_t \sim \mathcal{N}(0, 1)$$

$$\Rightarrow y_t \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \lambda_i \underbrace{\left(\frac{1}{1 - \delta/2\lambda_i}\right)}\right)$$

slightly too big ☹
but mainly affects λ_{min},
not λ_{max} ☺

Also,

$$\operatorname{Corr}\left[f(y_0), f(y_s) \right] \leq \left| 1 - \frac{\delta}{\lambda_i} \right|^s$$

$\Rightarrow$ To get the best convergence rate for sampling a $\mathcal{N}(0, \Sigma)$ distribution, we need to set

$$1 - \frac{\delta}{\lambda_{max}} = \frac{\delta}{\lambda_{min}} - 1$$

$$\Rightarrow \delta = \frac{2}{1/\lambda_{min} + 1/\lambda_{max}}$$

Bad conditioning sucks ☹

Resulting convergence rate is

$$\operatorname{Corr}\left[f(x_0), f(x_s) \right] \leq \left(\frac{K - 1}{K + 1} \right)^s$$

where $K = \dfrac{\lambda_{max}}{\lambda_{min}}$ is the condition number.

MALA (Metropolis - adjusted Langevin)

1. Sample $\xi_t \sim \mathcal{N}(0,1)$

2. Set $x'_{t+1} = x_t + \delta \cdot \nabla \log \pi(x_t) + \sqrt{2\delta} \cdot \xi_t$

3. Sample $u_t \sim \mathrm{Unif}(0,1)$

4. Accept if $u_t \leq p_t(\text{accept})$

where
$$p_t(\text{accept}) = \frac{\pi(x'_{t+1})}{\pi(x_t)} \cdot \frac{\exp\left(-\frac{1}{2}\left[\frac{-\Delta x - \delta \nabla \log \pi(x'_{t+1})}{\sqrt{2\delta}}\right]^2\right)}{\exp\left(-\frac{1}{2}\left[\frac{\Delta x - \delta \nabla \log \pi(x_t)}{\sqrt{2\delta}}\right]^2\right)}$$

$$= \frac{\pi(x'_{t+1})}{\pi(x_t)} \cdot \exp\left(\frac{1}{\delta}\left[\Delta x - \delta \nabla \log \pi(x_t)\right]^2 - \frac{1}{\delta}\left[\Delta x + \delta \nabla \log \pi(x'_{t+1})\right]^2\right)$$

This last step is annoying and doesn't make much of a difference, considering other areas. My Ph.D advisor always excludes it.

Hybrid Monte Carlo

$v \sim \mathcal{N}(0,1)$

For $i = 1$ to L:

$$v \leftarrow v + \frac{\delta}{2} \nabla \log \pi(x)$$

$$x \leftarrow x + \delta v$$

$$v \leftarrow v + \frac{\delta}{2} \nabla \log \pi(x)$$

Samples $\pi(x) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\|v\|^2\right)$

$$\begin{cases} dx = v\,dt \\ dv = \nabla \log \pi(x)\, dt \end{cases}$$

Conserves $H(x,v) = \frac{1}{2}\|v\|^2 - \log \pi(x)$

$$\frac{d}{dt} H(x,v) = v^T \nabla \log \pi(x) - v^T \nabla \log \pi(x) = 0$$

$\Rightarrow$ Like MALA, we go upstream following $\nabla \log \pi(x)$

$\Rightarrow$ Unlike MALA, now we have momentum!