Check your presentation slot
7 minutes hard limit
3 minutes Q & A (please participate)

Metropolis - Hastings
(MCMC)

The target is sampled by a Markov chain

→ $X_0 \sim \pi_0$

→ For $t = 1, 2 \ldots$

$$\tilde{X}_t \sim q(\cdot \mid X_t)$$
(proposal distribution)

$$X_{t+1} = \begin{cases} \tilde{X}_t & \text{with prob } \alpha(X_t, \tilde{X}_t) \\ X_t & \text{with prob } 1 - \alpha(X_t, \tilde{X}_t) \end{cases}$$

Acceptance ratio
$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)}{\pi(x)} \cdot \frac{q(x|y)}{q(y|x)}\right\}$$

- The stationary distribution $\{X_t\}$ converges to π

- $\mathrm{Unif}(X_1, \ldots, X_T) = \hat{\pi}_T$ (empirical measure on chain up to T)
$\xrightarrow[TV]{T\to\infty} \pi$

$$d_{TV}(\hat{\pi}_T, \pi) \xrightarrow{T\to\infty} 0$$

- $K(x, y)$: Prob. of transitioning from x to y
$$\mathbb{P}(X_{t+1} = y \mid X_t = x)$$

- π is invariant for $\{X_t\}$ if
$$\pi(A) = \int K(x, A)\, \pi(x)\, dx$$

- Detailed balance
$$\pi(x) K(x, y) = \pi(y) K(y, x)$$

- Goal: Target π is invariant for M-H Markov chain.

$$K(x, y) = q(y|x) \times \alpha(x, y)$$

Det. bal.
- $q(y|x)\, \alpha(x, y) \times \pi(x) \leftarrow$
$$= q(x|y)\, \alpha(y, x) \times \pi(y)$$

- $$\alpha(x, y) = \min\left\{1, \frac{\pi(x)}{\pi(y)} \frac{q(y|x)}{q(x|y)}\right\}$$
Set $\alpha(x, y) = 1$ or $\alpha(y, x) = 1$

→ Mixing time can (in general) increases exponentially with d (dimension of sample space)
$$d_{TV}(\hat{\pi}_T, \pi) \sim O(e^{-\lambda T})$$

→ acceptance ratio may be small
→ Can produce correlated samples

Metropolis-adjusted Langevin dynamics (MALA)

$$dX_t = \underbrace{\nabla \log \pi(X_t)}_{\text{drift / score of invariant measure}} + \sqrt{2}\, dW_t$$

(Overdamped)

$$X_t \sim \pi$$

→ Convergence is slow

→ Numerical integrators for SDEs
Euler-Maruyama scheme

MH

- Propose samples by simulating LD.
- Accept/reject

- Hamiltonian Dynamics (Deterministic) + M-H

Hamiltonian Monte Carlo

Langevin dynamics

$$p(\dot{X}, X) \propto e^{-\beta \frac{\dot{X}^2}{2m}} e^{-\beta U(X)}$$

β : inverse temperature

$$d\ddot{X} = \underline{f(X(t))}\,dt - \gamma \dot{X}(t)\,dt + \sqrt{\frac{2m\gamma}{\beta}}\, dW(t)$$

$$\boxed{f(X) = -\frac{\partial U}{\partial X}}$$

$\nabla \log p$: score

→ $dW(t)$: differential of a Wiener process / Brownian motion

- $\mathbb{E}[dW(t)] = 0$
- $\mathrm{Cov}(dW(t), dW(s)) = 0$
- $\mathbb{E}[dW^2(t)] = dt$

- Sample from π / Variational inference
  - (model for target π)
  - parameter optimization by ELBO maximization
  - Approximation error

MCMC
- guarantees for sampling from π
- particle / sample evolutions whose marginals converge to π
- Slow in high dimensions and do not effectively tackle multi-modality

↳ Optimal transport methods

↳ GANs

Generative modeling

$\cancel{\nabla \log \pi}$

$X_1, \dots, X_m \sim \pi$
(m samples)

$$\hat{\pi}_m = \mathrm{Unif}(X_1, \dots, X_m)$$

Want: Samples from π

Score-Generative modeling
Diffusion model

$\rightarrow \quad dY_t = f(Y_t)\,dt + \sigma\,dW_t$ (forward)

$\cdot\, Y_0 \sim \hat{\pi}_m$ (empirical measure approximating target π)

$\rightarrow \quad dX_t = -f(X_t)\,dt + \sigma\sigma^T \nabla \log p_{T-t}(X_t)\,dt + \sigma\,dB_t$ (Reverse)

$X_0 \sim \varrho_0$ (Gaussian)

Forward-backward SDEs

- Run forward SDE for a finite time T.
- Reverse (generating process) start with P_T (marginal of Y_T)
- Reverse process uses "scores" associated with P_t (marginal of Y_t)
- Marginal of X_T = Marginal of Y_0

Approximations

Need	have
• π	$\hat{\pi}_m$
• $\nabla \log P_t(x,t)$ $t \in [0,T]$	$S_\omega(x,t)$ (NN)
• Continuous SDE	Time integrator
• P_T	ϱ_0 (Gaussian)

As a result, $X_T \sim \hat{\pi}$ (SGM)

$s_\omega(x,t)$: ERM denoising Score Matching

$$\ell(\omega) = \int_0^T \alpha(t)\, \mathbb{E}_{P_t} \| s_\omega(x,t) - \nabla \log P_t(x) \|\, dt$$