Sampling Method

Rejection Sampling

Goal

Sampling $X \sim p(x)$.

Algorithm

1. Generate $X \sim q(x)$
2. Accept X with probability $\frac{p(x)}{Mq(x)}$

Disccusion

- We require $Mq(x)$ cover $p(x)$
- The acceptance rate is $1/M$.

Matropolis-Hasting Algorithm

Goal

The target density is $\pi(x)$, for which we know its unnormalized density function $\pi_u(x)$.

Algorithm

1. Generate $Y_t \sim q\left(y \mid X^{(t)}\right)$
2. Compute $\rho\left(X^{(t)}, Y_t\right)$ with $\rho(x, y) = \min\left(1, \frac{\pi_u(y) q(x|y)}{\pi_u(x) q(y|x)}\right)$
3. Take

$$X^{(t+1)} = \begin{cases} Y_t & \text{with prob } \rho\left(X^{(t)}, Y_t\right) \\ X^{(t)} & \text{with prob } 1 - \rho\left(X^{(t)}, Y_t\right) \end{cases}$$

Variants

- **Symmetric Metropolis Algorithm**: $q(y \mid x) = q(x \mid y)$;
- **Random Walk Metropolis-Hastings**: $q(y \mid x) = q(y - x)$;
- **Indepencence Sampler**: $q(y \mid x) = q(y)$;
- **Langevin algrithm**: $q(y \mid x) = N\left(X^t + (\delta/2)\nabla \log \pi\left(X^t\right), \delta\right)$

Discussion

- The transition kernel is:

$$k(x, y)=\underbrace{q(y \mid x) \rho(x, y)}_{\text{Propose y and accept}}+\underbrace{(1-r(x)) \delta_x(y)}_{\text{Reject, stay at } x}$$

  where $r(x):=P(\text{ Accept })=\int q(y \mid x) \rho(x, y) d y$, the probability that new generated Y_{t+1} gets accepted.

- The MH chain satisfies the detailed balanced condition $K(y, x) \pi(y)=K(x, y) \pi(x)$.

- Independence sampler requires $\pi(y) \leq M q(y)$, the acceptance probability is at least $1 / M$ when stationary and MC is uniformly ergodic

$$\left\|K^n(x, \cdot)-\pi\right\|_{T V} \leq 2\left(1-\frac{1}{M}\right)^n$$

Langenvin Algorithm

Goal

Finding the maximum of a positive function, in some sense, is equivalent to finding the mode of a probability function without knowing constant. So, we can consider the question from the opposite direction, that is, finding the "distribution" of the positive function.

Here, consider the Langevin diffusion L_t defined by the stochastic differential equation

$$d L_t=d B_t+\frac{1}{2} \nabla \log \pi\left(L_t\right) d t$$

where B_t is a standard Brownian motion

Algorithm

1. Generate next Y_t by $Y_t=X^{(t)}+\frac{\sigma^2}{2} \nabla \log \pi_u\left(X^{(t)}\right)+\sigma \epsilon_t$, where ϵ_t is from standard Gaussion distribution.
2. Compute $\rho\left(X^{(t)}, Y_t\right)$ with $\rho(x, y)=\min \{1, \gamma(x, y)\}$ with

$$\gamma(x, y)=\frac{\pi_u(y) \times \exp \left\{-\left\|x-y-\frac{\sigma^2}{2} \nabla \log \pi_u(y)\right\|^2 /\left(2 \sigma^2\right)\right\}}{\pi_u(x) \times \exp \left\{-\left\|y-x-\frac{\sigma^2}{2} \nabla \log \pi_u(x)\right\|^2 /\left(2 \sigma^2\right)\right\}}$$

3. Take

$$X^{(t+1)}=\begin{cases}Y_t & \text{with prob } \rho\left(X^{(t)}, Y_t\right) \\ X^{(t)} & \text{with prob } 1-\rho\left(X^{(t)}, Y_t\right)\end{cases}$$

Disccusion

- The isotropic diffusion might be ineffcient for strongly correlated variables. The issue can be circumvented by employing a preconditioning matrix M such that

$$Y_t = X^{(t)} + \frac{\sigma^2}{2} \mathrm{M} \nabla \log \pi_u \left(X^{(t)} \right) + \sigma \mathrm{M}^{1/2} \epsilon_t$$

Hamiltonian MCMC

Goal

Target distribution $\pi(x) \propto \exp(-U(x))$, $U(x)$ is a potential. Augment the parameter space $\mathcal{X}$ to $\mathcal{X} \times \mathcal{V}$ and the target distribution $\pi(x)$ to the *canonical distribution* $\pi(x, v) = \pi(v \mid x)\pi(x)$.

For simplicity $\pi(v \mid x) \propto \exp\left(-v^T \mathbf{M}^{-1} v\right)$. Define the Hamiltionian as $H(x, v) = -\log \pi(x, v)$.

To summary, we want to sample $x \sim \pi$, but we sample (x, v) firstly and drop v. Then, the marginal is still the target distribution.

Algorithm

1. Generate next $Y_t = X^{(t)} - \frac{\epsilon^2}{2} \frac{\partial U}{\partial x} \left(X^{(t)} \right) + \epsilon \mathbf{M}^{-1} v^{(t)}$
2. Accept it with the probability $\min\left(1, \exp(-H\left(x', v'\right) + H(x, v)\right))$

Stochastic HMC

Goal

The Hamiltonian MCMC has acceptance procedure. With the idea of stochastic HMC, there is no need to include the acceptance procedure.

Consider a generic stochastic dynamics

$$dx = \mu(x) dt + \sqrt{2} \sigma(x) dB_t$$

Algorithm

The following stochastic gradient HMC with friction has proper stationary distribution

$$\begin{aligned} dx &= \mathrm{M}^{-1} v dt \\ dv &= -\nabla U(x) dt - \mathrm{B} \mathrm{M}^{-1} v dt + (2\mathrm{B})^{1/2} dB_t \end{aligned}$$