Hamiltonian Monte Carlo

(a.k.a. Hybrid Monte Carlo)

Kuhwan Jeong [1]

[1]Department of Statistics, Seoul National University, South Korea

January, 2017

Introduction

- HMC is a MCMC method that adopts physical system dynamics rather than a probability distribution to propose future states in the Markov chain.

- This allows the Markov chain to explore the target distribution much more efficiently, resulting in faster convergence.

Hamilton's Equations

- Let $x(t)$ be a location and $p(t)$ be a momentum at time t.
- For each location the object takes potential energy $U(x)$.
- For each momentum there is associated kinetic energy $K(p)$.
- The total energy of the system is constant and known as the Hamiltonian

$$H(x, p) = U(x) + K(p).$$

- The time evolution of the system is uniquely defined by Hamilton's equations:

$$\begin{aligned} \frac{dp}{dt} &= -\frac{\partial H}{\partial x} = -\frac{\partial U(x)}{\partial x}, \\ \frac{dx}{dt} &= \frac{\partial H}{\partial p} = \frac{\partial K(p)}{\partial p}. \end{aligned}$$

Discretizing Hamilton's Equations

- The Hamiltonian equations describe an object's motion in time.
- In order to simulate Hamiltonian dynamics numerically on a computer, it is necessary to approximate the Hamiltonian equations by discretizing time.

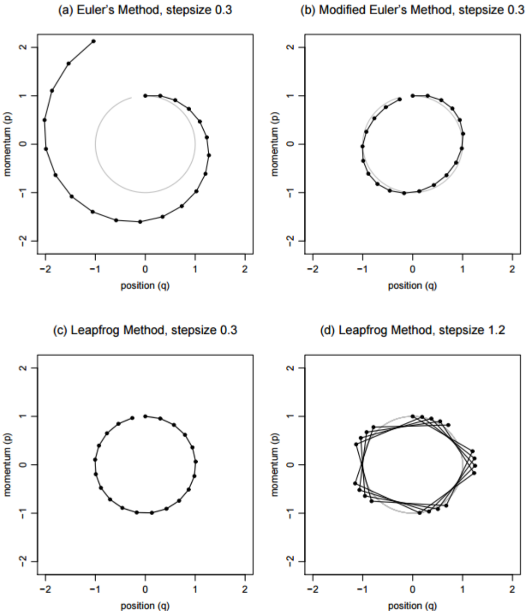
(1) Euler's Method

$$x_i(t+\delta) = x_i(t) + \delta \frac{dx_i}{dt}(t) = x_i(t) + \delta \frac{\partial K}{\partial p_i}(p(t)),$$
$$p_i(t+\delta) = p_i(t) + \delta \frac{dp_i}{dt}(t) = p_i(t) - \delta \frac{\partial U}{\partial x_i}(x(t)).$$

(2) The Leapfrog Method

$$p_i(t+\delta/2) = p_i(t) - (\delta/2) \frac{\partial U}{\partial x_i}(x(t)),$$
$$x_i(t+\delta) = x_i(t) + \delta \frac{\partial K}{\partial p_i}(p(t+\delta/2)),$$
$$p_i(t+\delta) = p_i(t+\delta/2) - (\delta/2) \frac{\partial U}{\partial x_i}(x(t+\delta)).$$

Discretizing Hamilton's Equations



Hamiltonian Monte Carlo

- We wish to sample d-dimensional x from

$$P(x) = \frac{1}{Z} \exp(-E(x)).$$

- We introduce a d-dimensional auxiliary variable p such that

$$p \sim \mathcal{N}(\mathbf{0}, M)$$

  where M is a symmetric, positive-definite matrix.

- The joint distribution of x and p is proportional to

$$\begin{aligned} P(x,p) &\propto \exp(-E(x)) \exp\big(-p^T M^{-1} p\big) \\ &= \exp\big(-E(x) - p^T M^{-1} p\big). \end{aligned}$$

- We define Hamiltonian function $H(x,p)$ and the kinetic energy $K(p)$ as

$$H(x,p) = E(x) + K(p), K(p) = p^T M^{-1} p/2.$$

Hamiltonian Monte Carlo

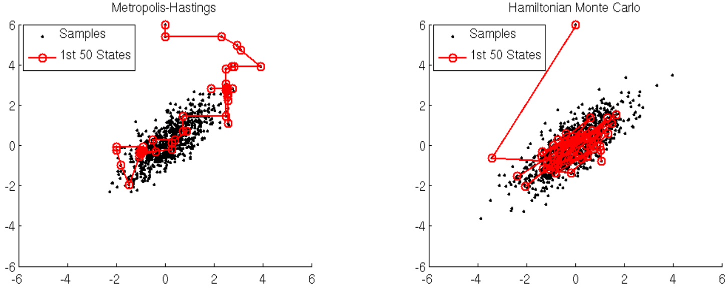
1. Set $m = 0$.
2. Generate an initial position $x^{(0)}$.
3. Repeat until $m = M$:
    1. Set $m = m + 1$.
    2. Sample $p_0 \sim \mathcal{N}(\mathbf{0}, M)$.
    3. Set $x_0 = x^{(m-1)}$.
    4. Starting from (x_0, p_0), do Leapfrog updates for L steps with stepsize δ to obtain (x^*, p^*).
    5. Calculate the Metropolis acceptance probability :

$$\alpha = \min\Bigg(1, \exp\big(-E(x^*) - K(p^*) + E(x_0) + K(p_0)\big)\Bigg).$$

    6. Sample $u \sim U(0, 1)$:
        - If $u \leq \alpha$, set $x^m = x^*$.
        - Otherwise, set $x^m = x^{m-1}$.

Simulation : Bivariate Normal distribution

- Bivariate normal distribution with $\rho = 0.8$.



Simulation : Bivariate Normal distribution

- 100-dimensional

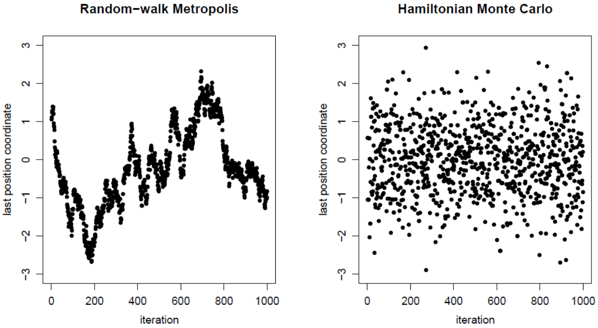


Figure 6: Values for the variable with largest standard deviation for the 100-dimensional example, from a random-walk Metropolis run and an HMC run with $L = 150$. To match computation time, 150 updates were counted as one iteration for random-walk Metropolis.

Reference

- https://theclevermachine.wordpress.com/2012/11/18/mcmc-hamiltonian-monte-carlo-a-k-a-hybrid-monte-carlo/

- Neal, Radford M. "MCMC using Hamiltonian dynamics." Handbook of Markov Chain Monte Carlo 2 (2011): 113-162.