Columbia University
IEOR4703 – Monte Carlo Simulation (Hirsa)
Assignment 6 – Due on Monday May 2nd, 2022 by midnight

Problem 1 (Hamiltonian Monte-Carlo): As you recall **Gibbs** Sampler failed to sample from the following distribution:

$$f(x_1, x_2) = \begin{cases} \frac{1}{2}: & 0 < x_1 < 1 \quad \& \quad 0 < x_2 < 1 \\ \frac{1}{2}: & -1 < x_1 < 0 \quad \& \quad -1 < x_2 < 0 \end{cases}$$

We can write $p(x) \propto \exp(0)\mathbb{1}_{\{x\in\mathcal{A}\}}$ where $\mathcal{A} = ([0,1]\times[0,1])\cup([-1,0]\times[-1,0])$. For Hamiltonian Monte Carlo, introduce the `momentum` variable y and we assume that it is normal, so that $p(y) \propto \exp(-\frac{1}{2}(y_1^2 + y_2^2))$. The joint density is then

$$p(x, y) = \frac{1}{Z}\exp\left(0 - \frac{1}{2}(y_1^2 + y_2^2)\right)\mathbb{1}_{\{x\in\mathcal{A}\}},$$

where Z is a constant. Form here, write Hamiltonian equations, use leapfrog discretization and the rest should be clear.

Problem 2 (Hamiltonian vs. Metropolis-Hastings): Consider a 10-dimensional multivariate normal distribution $\mathcal{N}(\theta; \mu, \Sigma)$ its pdf is given by

$$\pi(\theta) \propto \frac{1}{\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\theta-\mu)^\top \Sigma^{-1}(\theta-\mu)}$$

where $\theta = (\theta_1, \ldots \theta_{10})$. Assume $\mu = (0, \ldots, 0)$ and

$$\Lambda = \begin{pmatrix} 1 & \rho & \ldots & 0 \\ \rho & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix} \qquad \sigma = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 0.9 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0.2 & 0 \\ 0 & 0 & \ldots & \ldots & 0.1 \end{pmatrix}$$

where

$$\Sigma = \sigma^\top \Lambda\, \sigma$$

- In Metropolis-Hastings, assume the following distribution for proposal distribution, $q(\theta|\theta')$,

  $$\theta' = \theta + \eta Z \qquad Z \sim \mathcal{N}(0, I)$$

  Find Metropolis ratio and use Metropolis-Hasting algorithm to sample from the 10-dimensional multivariate normal

- For Hamiltonian Monte Carlo, using the following kinetic energy[1]

  $$K(p) = p^\top p/2$$

  to find Hamiltonian equations and sample from the 10-dimensional multivariate normal.

Plot & Compare traceplots and running averages and conclude on each method's performance. Use few different values for ρ, e.g. 0.1, 0.5, .99, -0.1, -0.5, -0.99.

[1] where $p = (p_1, \ldots, p_d)$

IEOR 4703 - Monte Carlo Simulation Methods
Solutions for Assignment 8

Ali Hirsa/Weilong Fu

Spring 2022

Problem 1 (Hamiltonian Monte Carlo)**.** We wish to sample from the bivariate distribution $p(x) = \frac{1}{2}\mathbb{1}_{\{x\in A\}}$, where $A = ([0,1]\times[0,1]) \cup ([-1,0]\times[-1,0])$. Observe that we can write $p(x) \propto \exp(0)\mathbb{1}_{\{x\in A\}}$. To use Hamiltonian Monte Carlo, we introduce the *momentum* variable y and we assume that it is normal, so that $p(y) \propto \exp(-\frac{1}{2}(y_1^2+y_2^2))$. The joint density is then

$$p(x,y) = \frac{1}{Z}\exp\left(0 - \frac{1}{2}(y_1^2+y_2^2)\right)\mathbb{1}_{\{x\in A\}},$$

where Z is a constant. In this simple case, the Hamiltonian is $H(x,y) = \frac{1}{2}(y_1^2+y_2^2)$ and therefore, the Hamilton equations become

$$\begin{aligned}\Delta y &= \epsilon\nabla_x = 0\\ \Delta x &= -\epsilon\nabla_y = -\epsilon y\end{aligned}$$

.

From the above equations we observe that the momentum y will not change while using the Hamiltonian dynamics. This implies, in particular, that both the modified Euler and Leapfrog methods reduce to the Euler method. Given a point $x_n \in A$, $\epsilon > 0$ and a number of steps $\ell > 0$, the proposal is obtained using this algorithm:

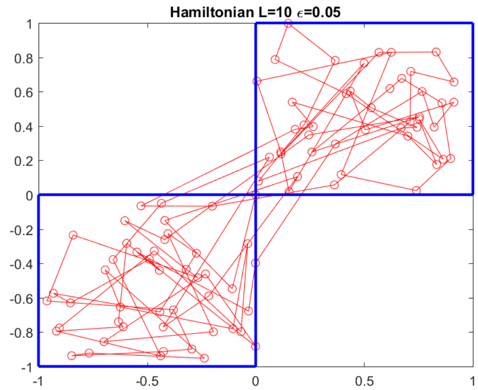
1. Simulate $y_1, y_2 \sim \mathrm{N}(0,1)$. $y_n = (y_1, y_2)$.
2. Set $x_{n+1} = x_n - \epsilon\ell y$ and $y_{n+1} = y_n$.

The acceptance of the proposal also simplifies considerably since

$$\frac{p(x_{n+1}, y_{n+1})}{p(x_n, y_n)} = \frac{p(x_{n+1})}{p(x_n)} = \mathbb{1}_{\{x_{n+1}\in A\}}.$$

Therefore, we accept x_{n+1} if it belongs to A and if not we set $x_{n+1} = x_n$. An implementation in Python using $\epsilon = 0.05$, $L = 10$ and 200 iterations produced the chain observed in Figure 1. Observe that using Hamiltonian Monte Carlo we are able to explore the two sub-regions of A.

Figure 1: Sample Markov chain using Hamiltonian Monte Carlo



Problem 2 (Hamiltonian vs. Metropolis-Hastings)**.** The Hamiltonian is $H(\theta, p) = \frac{1}{2}\theta^\top \Sigma^{-1}\theta + \frac{1}{2}p^\top p$ and the Hamilton equations become

$$\begin{aligned}
\frac{\partial \theta}{\partial t} &= \frac{\partial H}{\partial p} = p \\
\frac{\partial p}{\partial t} &= -\frac{\partial H}{\partial \theta} = -\Sigma^{-1}\theta
\end{aligned}$$

Leapfrog discretization leads to

$$\begin{aligned}
p(t + \frac{\varepsilon}{2}) &= p(t) - \frac{\varepsilon}{2}\Sigma^{-1}\theta(t) \\
\theta(t + \varepsilon) &= q\theta(t) + \varepsilon p(t + \frac{\varepsilon}{2}) \\
p(t + \varepsilon) &= p(t + \frac{\varepsilon}{2}) - \frac{\varepsilon}{2}\Sigma^{-1}\theta(t + \varepsilon)
\end{aligned}$$

Algorithm 1: The `HMC` algorithm

start with $\theta^{(0)}$
for $i = 1, \ldots, n$ **do**
  set $\theta = \theta^{(i-1)}$
  $p \sim \mathcal{N}(0, I)$ same size as θ
  set $p^{(i-1)} = p$
  $p = p - \frac{\varepsilon}{2}\Sigma^{-1}\theta$
  for $j = 1, \ldots, \ell - 1$ **do**
    $\theta = \theta + \varepsilon p$
    $p = p - \varepsilon\Sigma^{-1}\theta$
  end
  $\theta = \theta + \varepsilon p$
  $p = p - \frac{\varepsilon}{2}\Sigma^{-1}\theta$
  $H = \frac{1}{2}\theta^{(i-1)}\Sigma^{-1}\theta^{(i-1)} + \frac{1}{2}{p^{(i-1)}}^{\top} p^{(i-1)}$
  $\tilde{H} = \frac{1}{2}\theta^{\top}\Sigma^{-1}\theta + \frac{1}{2}p^{\top}p$
  $U \sim \mathcal{U}(0, 1)$
  if $U < e^{H - \tilde{H}}$ **then**
    $\theta^{(i)} = \theta$
  else
    $\theta^{(i)} = \theta^{(i-1)}$
  end
end

Metropolis-Hastings algorithm is straightforward.