Problem: random walk

Slow mixing due to “random walk” behavior

Why? Typical proposal is a random displacement
- Spherical Gaussian → Brownian motion-like
- Ignores density surface ($P$)

Review: Metropolis-Hastings

- Given: probability density $P(x)$, $x \in \mathbb{R}^d$
- Goal: generate sample $x \sim P$

Metropolis-Hastings

- Initialize $x^{(0)}$ arbitrarily
- Given $x^{(t)} = x$, propose
  $$x' \sim Q(\cdot \mid x)$$
- Accept and set $x^{(t+1)} = x'$ with probability $\min(a, 1)$
  $$a = \frac{P(x')}{P(x)} \cdot \frac{Q(x \mid x')}{Q(x' \mid x)}$$
- Else reject and set $x^{(t+1)} = x^{(t)}$

For large enough $T$, have $x^{(T)} \sim P$
Main Idea

- Idea: use density to guide proposals
- Select random velocity $\frac{p}{m} \in \mathbb{R}^d$
  - $p = \text{momentum}, \ m = \text{mass}$
- Simulate motion on energy surface

$$\{(x, E(x)) : x \in \mathbb{R}^d \} \subseteq \mathbb{R}^{d+1}, \ E(x) = -\log P(x)$$

with initial velocity $\frac{p}{m}$ for some amount of time to get proposal $x'$.

Hamiltonian Mechanics

- Position $x \in \mathbb{R}^d$
- Velocity $\frac{p}{m} \in \mathbb{R}^d$
- Potential energy $E(x)$ (= height)

- Temporal dynamics

$$\begin{align*}
\frac{dx}{dt} &= \frac{p}{m} \\
\frac{dp}{dt} &= -\frac{\partial E(x)}{\partial x}
\end{align*}$$

Puck of mass $m$ sliding on frictionless surface with velocity $\frac{p}{m}$, height at $x$ equal to $E(x)$ (and thus "incline" $\partial E(x)/\partial x$).

Generalization: Kinetic Energy

Define $K(p) = \frac{p^TP}{2m}$

kinetic energy

$$\begin{align*}
\frac{dx}{dt} &= \frac{p}{m} \\
\frac{dp}{dt} &= -\frac{\partial E(x)}{\partial x}
\end{align*}$$

$$\begin{align*}
\frac{dx}{dt} &= \frac{\partial K(p)}{\partial p} \\
\frac{dp}{dt} &= -\frac{\partial E(x)}{\partial x}
\end{align*}$$
Generalization: The Hamiltonian

Define \( H(x, p) = E(x) + K(p) \)

- Hamiltonian or total energy
- Euler's method

\[
\frac{\partial}{\partial p} \left( E(x) + K(p) \right) = \frac{\partial}{\partial p} K(p)
\]

\[
\frac{\partial}{\partial x} H(x, p) = \frac{\partial}{\partial x} E(x)
\]

\[
\frac{dx}{dt} = \frac{\partial H(x, p)}{\partial p}
\]

\[
\frac{dp}{dt} = -\frac{\partial H(x, p)}{\partial x}
\]

Simulating Hamiltonian Mechanics

Euler's method

\[
x(t + \varepsilon) = x(t) + \varepsilon \frac{p(t)}{m}
\]

\[
p(t + \varepsilon) = p(t) - \varepsilon \frac{\partial E(x(t))}{\partial x}
\]

Problem: numerically unstable

Leapfrog Method

Euler

\[
p(t + \frac{\varepsilon}{2}) \rightarrow p(t + \varepsilon)
\]

\[
x(t + \varepsilon) \rightarrow x(t + \varepsilon)
\]

More accurate and stable method

\[
p(t + \varepsilon/2) = p(t) - (\varepsilon/2) \frac{\partial E(x(t))}{\partial x}
\]

\[
x(t + \varepsilon) = x(t) + \varepsilon \frac{p(t + \varepsilon/2)}{m}
\]

\[
p(t + \varepsilon) = p(t + \varepsilon/2) - (\varepsilon/2) \frac{\partial E(x(t + \varepsilon))}{\partial x}
\]
Leapfrog Method

Handbook of Markov Chain Monte Carlo

at times $\varepsilon, 2\varepsilon, 3\varepsilon, \ldots$, and hence find (approximate) values for $q(\tau)$ and $p(\tau)$ after $\tau/\varepsilon$ steps (assuming $\tau/\varepsilon$ is an integer).

Figure 5.1a shows the result of using Euler's method to approximate the dynamics defined by the Hamiltonian of Equation 5.8, starting from $q(0) = 0$ and $p(0) = 1$, and using a stepsize of $\varepsilon = 0.3$ for 20 steps (i.e. to $\tau = 0.3 \times 20 = 6$). The results are not good—Euler's method produces a trajectory that diverges to infinity, but the true trajectory is a circle. Using a smaller value of $\varepsilon$, and correspondingly more steps, produces a more accurate result at $\tau = 6$, but although the divergence to infinity is slower, it is not eliminated.

(a) Momentum ($p$)

(b) Modified Euler's method, stepsize 0.3

(c) Leapfrog method, stepsize 0.3

(d) Leapfrog method, stepsize 1.2

Hamiltonian MCMC

Random velocity/momentum instead of random displacement

- Start at $x$
- Choose random momentum $p \sim \exp(-p^T p / 2m)$
- Simulate Hamiltonian mechanics for $s$ time units to end at $x'$
- Propose $x'$

Problem: how to compute $Q(x' | x)$ for acceptance probability?

$$a = \frac{P(x')}{P(x)} \cdot \frac{Q(x | x')}{Q(x' | x)}$$

Auxilliary Variables

Sample both $x$ and $p$ from

$P(x, p) = \exp(-H(x, p)) = \exp(-E(x)) \exp(-K(p))$,

when done, discard $p$ values

Note: $x$ and $p$ are independent

Hamiltonian MCMC

Gibbs step

$(x, p) \rightarrow (x', p')$

$Q(x', p' | x, p)$

- Start at $(x, p)$
- Choose random momentum $p \sim \exp(-p^T p / 2m)$
- End at $(x, p)$

Metropolis-Hastings step

- Start at $(x, p)$
- Simulate Hamiltonian mechanics to end at $(x', p')$
- Propose $(x', -p')$
**Acceptance Probability?**

$\kappa(p) \propto \frac{P(x', -p')}{P(x, p)} \frac{Q(x, p | x', -p')}{Q(x', -p' | x, p)}$

$\alpha = \frac{P(x', -p')}{P(x, p)} \exp(-H(x', p) - K(p))$

$= \exp(\mathcal{H}(x) - \mathcal{H}(x') + K(p) - K(p')) \quad (K(p') = K(-p'))$

$\approx 1 \quad \text{(conservation of energy)}$

**Reversibility**

$\mathcal{Q}(x' - p' | x, p) = \mathcal{Q}(x, p | x', -p')$

$T_{L,\varepsilon}$

Let $T_{L,\varepsilon} : \mathbb{R}^{2d} \rightarrow \mathbb{R}^{2d}$ be simulation mapping with $L$ steps at time increment $\varepsilon$

$T_{L,\varepsilon}(x, p) = (x', p') \implies T_{L,\varepsilon}(x', -p') = (x, -p)$

**Demo**

Demo with sampling

**Example**

**Hamiltonian Monte Carlo**

(a) $\mathcal{H}(x) = E(x) + K(p)$

(b) $\mathcal{H}(x) = E(x) + K(p)$

(c) $\mathcal{H}(x) = E(x) + K(p)$

**Simple Metropolis**

(d) $\mathcal{H}(x) = E(x) + K(p)$

(e) $\mathcal{H}(x) = E(x) + K(p)$

(f) $\mathcal{H}(x) = E(x) + K(p)$
Example

Setup: 100D Gaussian, standard deviations in different dimensions are 0.01, 0.02, ..., 1.00

MCMC Using Hamiltonian Dynamics

With the same standard deviation. As discussed below in Section 5.4.1, the performance of both these sampling methods is invariant to rotation, so this example is illustrative of how they perform on any multivariate Gaussian distribution in which the square roots of the eigenvalues of the covariance matrix are 0.01, 0.02, ..., 0.99, 1.00.

For this problem, the position coordinates, $q_i$, and corresponding momentum coordinates, $p_i$, are all independent, so the leapfrog steps used to simulate a trajectory operate independently for each $(q_i, p_i)$ pair. However, whether the trajectory is accepted depends on the total error in the Hamiltonian due to the leapfrog discretization, which is a sum of the errors due to each $(q_i, p_i)$ pair (for the terms in the Hamiltonian involving this pair).

Keeping this error small requires limiting the leapfrog stepsize to a value roughly equal to the smallest of the standard deviations (0.01), which implies that many leapfrog steps will be needed to move a distance comparable to the largest of the standard deviations (1.00). Consistent with this, I applied HMC to this distribution using trajectories with $L = 150$ and with $\epsilon$ randomly selected for each iteration, uniformly from $(0.0104, 0.0156)$, which is $0.013 \pm 20\%$. I used random-walk Metropolis with proposal standard deviation drawn uniformly from $(0.0176, 0.0264)$, which is $0.022 \pm 20\%$. These are close to optimal settings for both methods. The rejection rate was 0.13 for HMC and 0.75 for random-walk Metropolis.

Figure 5.6 shows results from runs of 1000 iterations of HMC (right) and of random-walk Metropolis (left), counting 150 random-walk Metropolis updates as one iteration, so that the computation time per iteration is comparable to that for HMC. The plot shows the last variable, with the largest standard deviation. The autocorrelation of these values is clearly much higher for random-walk Metropolis than for HMC. Figure 5.7 shows the estimates for the mean and standard deviation of each of the 100 variables obtained using the HMC and random-walk Metropolis runs (estimates were just the sample means and sample standard deviations of the values from the 1000 iterations). Except for the first few