Hamiltonian MCMC

Dan Sheldon

April 15, 2024

Metropolis-Hastings

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

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

Review: Metropolis-Hastings

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

Problem: random walk

Slow mixing due to "random walk" behavior

- Typical proposal is a random displacement
  - Spherical Gaussian $\rightarrow$ Brownian motion-like
  - Ignores density surface

Why?

- Simple random-walk Metropolis method, given equal computer time.

Main Idea

▶ Idea: use density to guide proposals
▶ Select random velocity $p/m \in \mathbb{R}^d$
  ▶ $p =$ momentum, $m =$ mass
▶ Simulate motion on energy surface
  \[
  \{(x, E(x)) : x \in \mathbb{R}^d \} \subseteq \mathbb{R}^{d+1}, \quad E(x) = -\log P(x)
  \]
  with initial velocity $p/m$ for some amount of time to get proposal $x'$.

Hamiltonian Mechanics

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

▶ Temporal dynamics
  \[
  \frac{dx}{dt} = \frac{p}{m}, \quad \frac{dp}{dt} = -\frac{\partial E(x)}{\partial x}
  \]

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

Generalization: Kinetic Energy

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

kinetic energy

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

$\Rightarrow$

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

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

Hamiltonian or total energy

$\frac{dx}{dt} = \frac{\partial K(p)}{\partial p}$
$\frac{dp}{dt} = -\frac{\partial E(x)}{\partial x}$

$\Rightarrow$

Simulating Hamiltonian Mechanics

Euler's method

$x(t + \varepsilon) = x(t) + \varepsilon p(t)$

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

Problem: numerically unstable

Leapfrog Method

More accurate and stable method

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

$x(t + \varepsilon) = x(t) + \varepsilon p(t + \varepsilon/2)$

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

Figure 5.1 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 ($p$) Euler’s method, stepsize 0.3
(b) Modified Euler’s method, stepsize 0.3
(c) (d) Leapfrog method, stepsize 0.3

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 $\rightarrow$ end at $x'$
- Propose $x'$

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

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
- 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 $\rightarrow$ end at $(x', p')$
- Propose $(x', -p')$
Acceptance Probability?

\[
\begin{align*}
a &= \frac{P(x', -p')}{P(x, p)} \cdot \frac{Q(x, p | x', -p')}{Q(x', -p' | x, p)} \\
&= \frac{P(x', -p')}{P(x, p)} & \text{(reversibility, volume preservation)} \\
&= \exp(E(x) - E(x') + K(p) - K(p')) & (K(p') = K(-p')) \\
&\approx 1 & \text{(conservation of energy)}
\end{align*}
\]

Reversibility

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

\[ T_{L,\epsilon}(x, p) = (x', p') \implies T_{L,\epsilon}(x', -p') = (x, -p) \]

Demo

Demo with sampling

Example

Hamiltonian Monte Carlo

Simple Metropolis
Example

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

![Graphs showing random-walk Metropolis and Hamiltonian Monte Carlo](image)