STAT 576 Bayesian Analysis

Lecture 9: Hybrid Monte Carlo

Chencheng Cai

Washington State University

Hybrid Monte Carlo

- Hybrid Monte Carlo is a MCMC algorithm that combines the Metropolis algorithm with molecular dynamics (MD).
- The molecular dynamics (MD) is a **deterministic** algorithm that simulates the motion of particles in a physical system.

Hybrid Monte Carlo

- Hybrid Monte Carlo is a MCMC algorithm that combines the Metropolis algorithm with molecular dynamics (MD).
- The molecular dynamics (MD) is a **deterministic** algorithm that simulates the motion of particles in a physical system.
- We use the MD to propose a new state of the system, and then use the Metropolis algorithm to accept or reject the proposed state.
- Benefits:
  - It can explore the state space more efficiently than the random walk Metropolis.
  - It can handle high-dimensional state space.

Newton's Mechanics

Consider a single particle with mass m in a potential energy field $U(\boldsymbol{q})$.

- The particle's **position** is denoted by $\boldsymbol{q}$ and its **momentum** by $\boldsymbol{p}$. In **non-relativistic (NR)** mechanics, the momentum is given by

$$\boldsymbol{p} = m\boldsymbol{v},$$

where $\boldsymbol{v} = \dot{\boldsymbol{q}} := d\boldsymbol{q}/dt$ is the velocity.

Newton's Mechanics

Consider a single particle with mass m in a potential energy field $U(\boldsymbol{q})$.

- The particle's **position** is denoted by $\boldsymbol{q}$ and its **momentum** by $\boldsymbol{p}$. In **non-relativistic (NR)** mechanics, the momentum is given by

$$\boldsymbol{p} = m\boldsymbol{v},$$

  where $\boldsymbol{v} = \dot{\boldsymbol{q}} := d\boldsymbol{q}/dt$ is the velocity.

- The **kinetic energy** of the particle is

$$T(\boldsymbol{p}) = \frac{\|\boldsymbol{p}\|^2}{2m} \overset{\text{NR}}{=} \frac{1}{2}m\|\boldsymbol{v}\|^2.$$

Newton's Mechanics

The (NR) dynamics of this particle is determined by Newton's second law:

$$\boldsymbol{F} = m\boldsymbol{a},$$

where $\boldsymbol{F}$ is the force acting on the particle, and $\boldsymbol{a} = \ddot{\boldsymbol{q}} = d^2\boldsymbol{q}/dt^2$ is the acceleration.

Newton's Mechanics

The (NR) dynamics of this particle is determined by Newton's second law:

$$\boldsymbol{F} = m\boldsymbol{a},$$

where $\boldsymbol{F}$ is the force acting on the particle, and $\boldsymbol{a} = \ddot{\boldsymbol{q}} = d^2\boldsymbol{q}/dt^2$ is the acceleration.

- If the force is derived from a potential energy field $U(\boldsymbol{q})$, then the force is given by

$$\boldsymbol{F} = -\frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}}$$

Newton's Mechanics

The (NR) dynamics of this particle is determined by Newton's second law:

$$\boldsymbol{F} = m\boldsymbol{a},$$

where $\boldsymbol{F}$ is the force acting on the particle, and $\boldsymbol{a} = \ddot{\boldsymbol{q}} = d^2\boldsymbol{q}/dt^2$ is the acceleration.

- If the force is derived from a potential energy field $U(\boldsymbol{q})$, then the force is given by

$$\boldsymbol{F} = -\frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}}$$

- The $m\boldsymbol{a}$ is the time derivative of the momentum:

$$m\boldsymbol{a} = \dot{\boldsymbol{p}} = \frac{d\boldsymbol{p}}{dt}$$

Newton's Mechanics

The (NR) dynamics of this particle is determined by Newton's second law:

$$\boldsymbol{F} = m\boldsymbol{a},$$

where $\boldsymbol{F}$ is the force acting on the particle, and $\boldsymbol{a} = \ddot{\boldsymbol{q}} = d^2\boldsymbol{q}/dt^2$ is the acceleration.

- If the force is derived from a potential energy field $U(\boldsymbol{q})$, then the force is given by

$$\boldsymbol{F} = -\frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}}$$

- The $m\boldsymbol{a}$ is the time derivative of the momentum:

$$m\boldsymbol{a} = \dot{\boldsymbol{p}} = \frac{d\boldsymbol{p}}{dt}$$

- The Newton's second law in the **phase space** $(\boldsymbol{q}, \boldsymbol{p})$ is

$$-\frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}} = \dot{\boldsymbol{p}}.$$

Lagrangian Analytical Mechanics

We consider a (NR) particle in a system, whose dynamics is parametrized by $(t, \boldsymbol{q}, \dot{\boldsymbol{q}})$.

- We observe that

$$\frac{\partial T(\dot{\boldsymbol{q}})}{\partial \dot{\boldsymbol{q}}} = m\dot{\boldsymbol{q}}, \quad \boldsymbol{F} = -\frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}}$$

Lagrangian Analytical Mechanics

We consider a (NR) particle in a system, whose dynamics is parametrized by $(t, \boldsymbol{q}, \dot{\boldsymbol{q}})$.

- We observe that

$$\frac{\partial T(\dot{\boldsymbol{q}})}{\partial \dot{\boldsymbol{q}}} = m\dot{\boldsymbol{q}}, \quad \boldsymbol{F} = -\frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}}$$

- By Newton's second law, we have

$$\frac{d}{dt}\frac{\partial T(\dot{\boldsymbol{q}})}{\partial \dot{\boldsymbol{q}}} + \frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}} = 0$$

Lagrangian Analytical Mechanics

We consider a (NR) particle in a system, whose dynamics is parametrized by $(t, \boldsymbol{q}, \dot{\boldsymbol{q}})$.

- We observe that
$$\frac{\partial T(\dot{\boldsymbol{q}})}{\partial \dot{\boldsymbol{q}}} = m\dot{\boldsymbol{q}}, \quad \boldsymbol{F} = -\frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}}$$

- By Newton's second law, we have
$$\frac{d}{dt}\frac{\partial T(\dot{\boldsymbol{q}})}{\partial \dot{\boldsymbol{q}}} + \frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}} = 0$$

- If we define the **Lagrangian** of the system by
$$L(t, \boldsymbol{q}, \dot{\boldsymbol{q}}) = T(\dot{\boldsymbol{q}}) - U(\boldsymbol{q}),$$
then the equation above is equivalent to the **Euler-Lagrange equation**:
$$\frac{\partial L}{\partial \boldsymbol{q}} - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} = 0$$

Lagrangian Analytical Mechanics

We consider a particle in a system, whose dynamics is parametrized by $(t, \boldsymbol{q}, \boldsymbol{p})$.

- By Newton's second law, we have

$$-\frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}} = \dot{\boldsymbol{p}}.$$

Lagrangian Analytical Mechanics

We consider a particle in a system, whose dynamics is parametrized by $(t, \boldsymbol{q}, \boldsymbol{p})$.

- By Newton's second law, we have

$$-\frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}} = \dot{\boldsymbol{p}}.$$

- On the other hand, we have

$$\frac{\partial T(\boldsymbol{p})}{\partial \boldsymbol{p}} = \dot{\boldsymbol{q}}$$

Lagrangian Analytical Mechanics

We consider a particle in a system, whose dynamics is parametrized by $(t, \boldsymbol{q}, \boldsymbol{p})$.

- By Newton's second law, we have

$$-\frac{\partial U(\boldsymbol{q})}{\partial \boldsymbol{q}} = \dot{\boldsymbol{p}}.$$

- On the other hand, we have

$$\frac{\partial T(\boldsymbol{p})}{\partial \boldsymbol{p}} = \dot{\boldsymbol{q}}$$

- Combnining the two equations above, we have the **Hamilton's equations**:

$$\dot{\boldsymbol{q}} = \frac{\partial H(\boldsymbol{q}, \boldsymbol{p})}{\partial \boldsymbol{p}},$$
$$\dot{\boldsymbol{p}} = -\frac{\partial H(\boldsymbol{q}, \boldsymbol{p})}{\partial \boldsymbol{q}},$$

where $H(\boldsymbol{q}, \boldsymbol{p}) = T(\boldsymbol{p}) + U(\boldsymbol{q})$ is the **Hamiltonian** of the system.

Lagrangian Analytical Mechanics

The Hamiltonian's equations can be derived from the Euler-Lagrange equation directly.

- We define the momentum $\boldsymbol{p}$ as the conjugate variable of the position $\boldsymbol{q}$:

$$\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}$$

Lagrangian Analytical Mechanics

The Hamiltonian's equations can be derived from the Euler-Lagrange equation directly.

- We define the momentum $\boldsymbol{p}$ as the conjugate variable of the position $\boldsymbol{q}$:

$$\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}$$

- The Hamiltonian is defined as the Legendre transform of the Lagrangian:

$$H(\boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{p} \cdot \dot{\boldsymbol{q}} - L(\boldsymbol{q}, \dot{\boldsymbol{q}})$$

Lagrangian Analytical Mechanics

The Hamiltonian's equations can be derived from the Euler-Lagrange equation directly.

- We define the momentum $\boldsymbol{p}$ as the conjugate variable of the position $\boldsymbol{q}$:

$$\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{q}}}$$

- The Hamiltonian is defined as the Legendre transform of the Lagrangian:

$$H(\boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{p} \cdot \dot{\boldsymbol{q}} - L(\boldsymbol{q}, \dot{\boldsymbol{q}})$$

- The Hamiltonian's equations are derived from the Euler-Lagrange equation:

$$\frac{\partial H}{\partial \boldsymbol{p}} = \dot{\boldsymbol{q}},$$
$$\frac{\partial H}{\partial \boldsymbol{q}} = -\frac{\partial L}{\partial \boldsymbol{q}} = -\frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} = -\dot{\boldsymbol{p}}.$$

Molecular Dynamics Simulation

Consider simulating a system with Hamiltonian $H(\boldsymbol{q}, \boldsymbol{p})$.

- Use Taylor expansion:

$$\begin{aligned}
\boldsymbol{q}(t+dt) &= \boldsymbol{q}(t) + \dot{\boldsymbol{q}}(t)dt + \frac{1}{2}\ddot{\boldsymbol{q}}(t)(dt)^2 + \cdots, \\
&= \boldsymbol{q}(t) + \frac{\boldsymbol{p}(t)}{m}dt + \frac{1}{2}\frac{\dot{\boldsymbol{p}}(t)}{m}(dt)^2 + \cdots, \\
\boldsymbol{p}(t+dt) &= \boldsymbol{p}(t) + \dot{\boldsymbol{p}}(t)dt + \frac{1}{2}\ddot{\boldsymbol{p}}(t)(dt)^2 + \cdots.
\end{aligned}$$

Molecular Dynamics Simulation

Consider simulating a system with Hamiltonian $H(\boldsymbol{q}, \boldsymbol{p})$.

- Use Taylor expansion:

$$\begin{aligned}
\boldsymbol{q}(t+dt) &= \boldsymbol{q}(t) + \dot{\boldsymbol{q}}(t)dt + \frac{1}{2}\ddot{\boldsymbol{q}}(t)(dt)^2 + \cdots, \\
&= \boldsymbol{q}(t) + \frac{\boldsymbol{p}(t)}{m}dt + \frac{1}{2}\frac{\dot{\boldsymbol{p}}(t)}{m}(dt)^2 + \cdots, \\
\boldsymbol{p}(t+dt) &= \boldsymbol{p}(t) + \dot{\boldsymbol{p}}(t)dt + \frac{1}{2}\ddot{\boldsymbol{p}}(t)(dt)^2 + \cdots.
\end{aligned}$$

- Furthermore, we have

$$\begin{aligned}
\boldsymbol{q}(t+dt) + \boldsymbol{q}(t-dt) &= 2\boldsymbol{q}(t) + \frac{\dot{\boldsymbol{p}}(t)}{m}(dt)^2 + O((dt)^4), \\
\boldsymbol{q}(t+dt) - \boldsymbol{q}(t-dt) &= 2\boldsymbol{p}(t)dt + O((dt)^3).
\end{aligned}$$

Molecular Dynamics Simulation

Størmer-Verlet algorithm:

- At time t with a time increment Δt.
- We update the position by:

$$\boldsymbol{q}(t+\Delta t) = 2\boldsymbol{q}(t) - \boldsymbol{q}(t-\Delta t) - \frac{(\Delta t)^2}{m}\frac{\partial H}{\partial \boldsymbol{q}}\bigg|_t$$

- We update the momentum by:

$$\boldsymbol{p}(t+\Delta t) = m\frac{\boldsymbol{q}(t+\Delta t) - \boldsymbol{q}(t-\Delta t)}{2\Delta t}$$

or

$$\boldsymbol{p}(t+\Delta t) = m\frac{\boldsymbol{q}(t+\Delta t) - \boldsymbol{q}(t)}{\Delta t}$$

Molecular Dynamics Simulation

Leap-frog algorithm:

- At time t with a time increment Δt.
- We update the position by:

$$\boldsymbol{q}(t+\Delta t) = \boldsymbol{q}(t) + \Delta t \frac{\boldsymbol{p}\left(t+\frac{1}{2}\Delta t\right)}{m}$$

- We update the momentum at the half-time interval by:

$$\boldsymbol{p}\left(t+\frac{1}{2}\Delta t\right) = \boldsymbol{p}\left(t-\frac{1}{2}\Delta t\right) - \left.\frac{\partial H}{\partial \boldsymbol{q}}\right|_t \Delta t$$

Molecular Dynamics Simulation

Leap-frog algorithm (alternative form):

- At time t with a time increment Δt.
- We update the momentum by the first half-time interval:

$$\boldsymbol{p}\left(t+\frac{1}{2}\Delta t\right)=\boldsymbol{p}\left(t\right)-\left.\frac{\partial H}{\partial \boldsymbol{q}}\right|_{t}\frac{\Delta t}{2}$$

- We update the position by:

$$\boldsymbol{q}(t+\Delta t)=\boldsymbol{q}(t)+\Delta t\frac{\boldsymbol{p}\left(t+\frac{1}{2}\Delta t\right)}{m}$$

- We update the momentum by the second half-time interval:

$$\boldsymbol{p}\left(t+\Delta t\right)=\boldsymbol{p}\left(t+\frac{1}{2}\Delta t\right)-\left.\frac{\partial H}{\partial \boldsymbol{q}}\right|_{t+\Delta}\frac{\Delta t}{2}$$

Molecular Dynamics Simulation

Størmer-Verlet algorithm and Leap-frog algorithm are identical algorithms.

- From the position update set in leap-frog algorithm, we have

$$\boldsymbol{q}(t+\Delta t) = \boldsymbol{q}(t) + \Delta t \frac{\boldsymbol{p}\left(t+\frac{1}{2}\Delta t\right)}{m}$$

$$\boldsymbol{q}(t) = \boldsymbol{q}(t-\Delta t) + \Delta t \frac{\boldsymbol{p}\left(t-\frac{1}{2}\Delta t\right)}{m}$$

Molecular Dynamics Simulation

Størmer-Verlet algorithm and Leap-frog algorithm are identical algorithms.

- From the position update set in leap-frog algorithm, we have

$$\boldsymbol{q}(t+\Delta t) = \boldsymbol{q}(t) + \Delta t \frac{\boldsymbol{p}\left(t+\frac{1}{2}\Delta t\right)}{m}$$

$$\boldsymbol{q}(t) = \boldsymbol{q}(t-\Delta t) + \Delta t \frac{\boldsymbol{p}\left(t-\frac{1}{2}\Delta t\right)}{m}$$

- Substitute the second equation into the first equation, we have

$$\boldsymbol{q}(t+\Delta t) = 2\boldsymbol{q}(t) - \boldsymbol{q}(t-\Delta t) - \frac{(\Delta t)^2}{m}\frac{\partial H}{\partial \boldsymbol{q}}\bigg|_t$$

which is the same as the position update in Størmer-Verlet algorithm.

Molecular Dynamics Simulation

The Hamiltonian dynammics has the following preservation properties:

Molecular Dynamics Simulation

The Hamiltonian dynammics has the following preservation properties:

- The Hamiltonian $H(\boldsymbol{q}, \boldsymbol{p})$ is preserved by the dynamics.

$$\frac{dH}{dt} = \frac{\partial H}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}} + \frac{\partial H}{\partial \boldsymbol{p}}\dot{\boldsymbol{p}} = 0.$$

- The volume in the phase space is preserved by the dynamics.
  Let $V(t) = \{(\boldsymbol{q}(t), \boldsymbol{p}(t)) : (\boldsymbol{q}(0), \boldsymbol{p}(0)) \in V(0)\}$, then
  Louville's theorem: $V(t)$ has the same volume as $V(0)$.

Molecular Dynamics Simulation

The Hamiltonian dynammics has the following preservation properties:

- The Hamiltonian $H(\boldsymbol{q}, \boldsymbol{p})$ is preserved by the dynamics.

$$\frac{dH}{dt} = \frac{\partial H}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}} + \frac{\partial H}{\partial \boldsymbol{p}}\dot{\boldsymbol{p}} = 0.$$

- The volume in the phase space is preserved by the dynamics.
  Let $V(t) = \{(\boldsymbol{q}(t), \boldsymbol{p}(t)) : (\boldsymbol{q}(0), \boldsymbol{p}(0)) \in V(0)\}$, then
  Louville's theorem: $V(t)$ has the same volume as $V(0)$.

The MD simulation is volume-preservation and time reversable, but not Hamiltonian-preserving.

Molecular Dynamics Simulation

The Hamiltonian dynammics has the following preservation properties:

- The Hamiltonian $H(\boldsymbol{q}, \boldsymbol{p})$ is preserved by the dynamics.

$$\frac{dH}{dt} = \frac{\partial H}{\partial \boldsymbol{q}}\dot{\boldsymbol{q}} + \frac{\partial H}{\partial \boldsymbol{p}}\dot{\boldsymbol{p}} = 0.$$

- The volume in the phase space is preserved by the dynamics.
  Let $V(t) = \{(\boldsymbol{q}(t), \boldsymbol{p}(t)) : (\boldsymbol{q}(0), \boldsymbol{p}(0)) \in V(0)\}$, then
  Louville's theorem: $V(t)$ has the same volume as $V(0)$.

The MD simulation is volume-preservation and time reversable, but not Hamiltonian-preserving.

- The MD simulation is volume-preserving because the Jacobian of the transformation at each step is 1.
- The MD simulation is time reversable because we can simulate the system backward by reversing the momentum.
- Hamiltonian is not preserved because of the dicretization error.

Example

Consider a dynamic system with

$$U(q) = q^2 - 2\log[\cosh(2q)] + 3, \quad T(p) = p^2/2$$

Example

Consider a dynamic system with

$$U(q) = q^2 - 2\log[\cosh(2q)] + 3, \quad T(p) = p^2/2$$

The Hamiltonian is

$$H(q,p) = U(q) + T(p) = q^2 + \frac{p^2}{2} - 2\log[\cosh(2q)]$$

Example

Consider a dynamic system with

$$U(q) = q^2 - 2\log[\cosh(2q)] + 3, \quad T(p) = p^2/2$$

The Hamiltonian is

$$H(q,p) = U(q) + T(p) = q^2 + \frac{p^2}{2} - 2\log[\cosh(2q)]$$

The Hamiltonian's equations give

$$\dot{p} = -\frac{\partial H}{\partial q} = -2q + 4\tanh(2q)$$

Example

Consider a dynamic system with

$$U(q) = q^2 - 2\log[\cosh(2q)] + 3, \quad T(p) = p^2/2$$

The Hamiltonian is

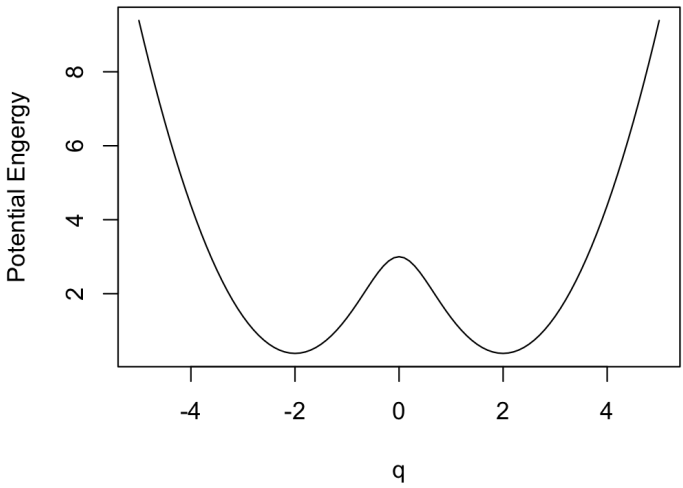
$$H(q,p) = U(q) + T(p) = q^2 + \frac{p^2}{2} - 2\log[\cosh(2q)]$$

The Hamiltonian's equations give

$$\dot{p} = -\frac{\partial H}{\partial q} = -2q + 4\tanh(2q)$$

We simulate the system using the leap-frog algorithm with $\Delta t = 0.1$, starting position $q(0) = 1$, and five different momentums $p(0) = 1, 2, 3, 4, 5$.

Example

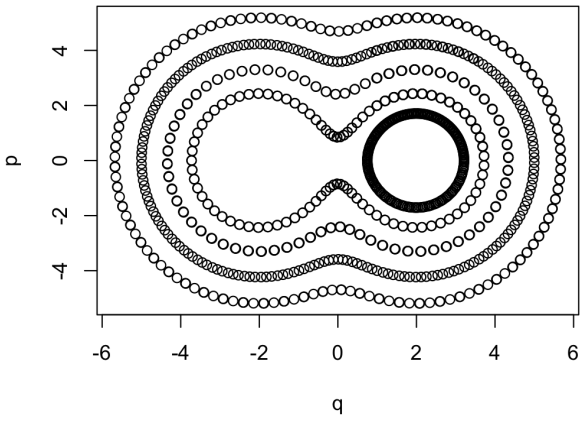


Example

```
n = 200
dt = 0.1

phase_sample = array(dim=c(5, 2, n+1))
phase_sample[,1,1] = 1
phase_sample[,2,1] = 1:5

pdot = function(q) {-2*q + 4*tanh(2*q)}

for(t in 1:n){
    p = phase_sample[,2,t] + 0.5 * pdot(phase_sample[,1,t]) * dt
    phase_sample[,1,t+1] = phase_sample[,1,t] + dt * p
    phase_sample[,2,t+1] = p + 0.5 * pdot(phase_sample[,1,t+1]) * dt
}
```

Example

Trajectories in the phase space of the system with different initial momentums.



Hamiltonian Monte Carlo

Suppose we have a target distribution $\pi(\boldsymbol{x}) \propto \exp\{-U(\boldsymbol{x})\}$, and we want to sample from it.

Hamiltonian Monte Carlo

Suppose we have a target distribution $\pi(\boldsymbol{x}) \propto \exp\{-U(\boldsymbol{x})\}$, and we want to sample from it.

- We introduce an auxiliary momentum variable $\boldsymbol{p}$ with density $p(\boldsymbol{p}) = \exp\{-T(\boldsymbol{p})\}$.

Hamiltonian Monte Carlo

Suppose we have a target distribution $\pi(\boldsymbol{x}) \propto \exp\{-U(\boldsymbol{x})\}$, and we want to sample from it.

- We introduce an auxiliary momentum variable $\boldsymbol{p}$ with density $p(\boldsymbol{p}) = \exp\{-T(\boldsymbol{p})\}$.
- The joint distribution of $(\boldsymbol{x}, \boldsymbol{p})$ is

$$p(\boldsymbol{x}, \boldsymbol{p}) = \exp\{-H(\boldsymbol{x}, \boldsymbol{p})\},$$

where $H(\boldsymbol{x}, \boldsymbol{p}) = U(\boldsymbol{x}) + T(\boldsymbol{p})$ is the Hamiltonian.

Hamiltonian Monte Carlo

Suppose we have a target distribution $\pi(\boldsymbol{x}) \propto \exp\{-U(\boldsymbol{x})\}$, and we want to sample from it.

- We introduce an auxiliary momentum variable $\boldsymbol{p}$ with density $p(\boldsymbol{p}) = \exp\{-T(\boldsymbol{p})\}$.
- The joint distribution of $(\boldsymbol{x}, \boldsymbol{p})$ is

$$p(\boldsymbol{x}, \boldsymbol{p}) = \exp\{-H(\boldsymbol{x}, \boldsymbol{p})\},$$

  where $H(\boldsymbol{x}, \boldsymbol{p}) = U(\boldsymbol{x}) + T(\boldsymbol{p})$ is the Hamiltonian.

Observations:

- If we can sample from the joint distribution $p(\boldsymbol{x}, \boldsymbol{p})$, then we can sample from the target distribution $\pi(\boldsymbol{x})$ by marginalizing out $\boldsymbol{p}$.

Hamiltonian Monte Carlo

Suppose we have a target distribution $\pi(\boldsymbol{x}) \propto \exp\{-U(\boldsymbol{x})\}$, and we want to sample from it.

- We introduce an auxiliary momentum variable $\boldsymbol{p}$ with density $p(\boldsymbol{p}) = \exp\{-T(\boldsymbol{p})\}$.
- The joint distribution of $(\boldsymbol{x}, \boldsymbol{p})$ is

$$p(\boldsymbol{x}, \boldsymbol{p}) = \exp\{-H(\boldsymbol{x}, \boldsymbol{p})\},$$

  where $H(\boldsymbol{x}, \boldsymbol{p}) = U(\boldsymbol{x}) + T(\boldsymbol{p})$ is the Hamiltonian.

Observations:

- If we can sample from the joint distribution $p(\boldsymbol{x}, \boldsymbol{p})$, then we can sample from the target distribution $\pi(\boldsymbol{x})$ by marginalizing out $\boldsymbol{p}$.
- The Hamiltonian dynamics is reversable, meaning that if $(\boldsymbol{x}(t), \boldsymbol{p}(t))$ results in $(\boldsymbol{x}(t + \Delta t), \boldsymbol{p}(t + \Delta t))$, then $(\boldsymbol{x}(t + \Delta t), -\boldsymbol{p}(t + \Delta t))$ results in $(\boldsymbol{x}(t), -\boldsymbol{p}(t))$.

Hamiltonian Monte Carlo

Suppose we have a target distribution $\pi(\boldsymbol{x}) \propto \exp\{-U(\boldsymbol{x})\}$. The Hamiltonian Monte Carlo algorithm is as follows: (for time $t+1$)

Hamiltonian Monte Carlo

Suppose we have a target distribution $\pi(\boldsymbol{x}) \propto \exp\{-U(\boldsymbol{x})\}$. The Hamiltonian Monte Carlo algorithm is as follows: (for time $t+1$)

- Sample a momentum $\boldsymbol{p}$ from

$$p(\boldsymbol{p}) \propto \exp\{-T(\boldsymbol{p})\}$$

- Run the leap-frog algorithm for L steps with step size ϵ jumping from $(\boldsymbol{x}, \boldsymbol{p})$ to $(\boldsymbol{x}', \boldsymbol{p}')$.
- Accept the new state $(\boldsymbol{x}', \boldsymbol{p}')$ with probability

$$\alpha = \min \left\{1, \exp\{-H(\boldsymbol{x}', \boldsymbol{p}') + H(\boldsymbol{x}, \boldsymbol{p})\}\right\}$$

Hamiltonian Monte Carlo

Suppose we have a target distribution $\pi(\boldsymbol{x}) \propto \exp\{-U(\boldsymbol{x})\}$. The Hamiltonian Monte Carlo algorithm is as follows: (for time $t+1$)

- Sample a momentum $\boldsymbol{p}$ from

$$p(\boldsymbol{p}) \propto \exp\{-T(\boldsymbol{p})\}$$

- Run the leap-frog algorithm for L steps with step size ϵ jumping from $(\boldsymbol{x}, \boldsymbol{p})$ to $(\boldsymbol{x}', \boldsymbol{p}')$.
- Accept the new state $(\boldsymbol{x}', \boldsymbol{p}')$ with probability

$$\alpha = \min \left\{1, \exp\{-H(\boldsymbol{x}', \boldsymbol{p}') + H(\boldsymbol{x}, \boldsymbol{p})\}\right\}$$

Remark: The leap-frog step can be replaced by any deterministic time-reversible and volume-preserving dynamics.

Hamiltonian Monte Carlo

Justification for correctness:

- (time-reversible) If the L leap-frog steps map $(\boldsymbol{x}, \boldsymbol{p})$ to $(\boldsymbol{x}', \boldsymbol{p}')$, then they also map $(\boldsymbol{x}', -\boldsymbol{p}')$ to $(\boldsymbol{x}, -\boldsymbol{p})$.
- (volume-preserving) The leap-frog algorithm is volume-preserving. Therefore $d\boldsymbol{x}d\boldsymbol{p} = d\boldsymbol{x}'d\boldsymbol{p}'$.

Hamiltonian Monte Carlo

Justification for correctness:

- (time-reversible) If the L leap-frog steps map $(\boldsymbol{x}, \boldsymbol{p})$ to $(\boldsymbol{x}', \boldsymbol{p}')$, then they also map $(\boldsymbol{x}', -\boldsymbol{p}')$ to $(\boldsymbol{x}, -\boldsymbol{p})$.
- (volume-preserving) The leap-frog algorithm is volume-preserving. Therefore $d\boldsymbol{x}d\boldsymbol{p} = d\boldsymbol{x}'d\boldsymbol{p}'$.
- The acceptance probability in a Metropolis-Hastings algorithm is

$$\alpha = \min\left\{1, \frac{\pi(\boldsymbol{x}')}{\pi(\boldsymbol{x})}\frac{p(\boldsymbol{x}', \boldsymbol{x})d\boldsymbol{x}'d\boldsymbol{p}'}{p(\boldsymbol{x}, \boldsymbol{x}')d\boldsymbol{x}d\boldsymbol{p}}\right\} = \min\left\{1, \frac{\pi(\boldsymbol{x}')}{\pi(\boldsymbol{x})}\frac{p(\boldsymbol{p}')}{p(\boldsymbol{p})}\right\} = \min\left\{1, \frac{e^{-H(\boldsymbol{x}', \boldsymbol{p})}}{e^{-H(\boldsymbol{x}, \boldsymbol{p})}}\right\}$$

Example

Consider sampling from the $\mathrm{Beta}(2, 2)$ distribution. The Hamiltonian is

$$H(x, p) = -\log(x(1 - x)) + \frac{p^2}{2}$$

Example

Consider sampling from the $\mathrm{Beta}(2, 2)$ distribution. The Hamiltonian is

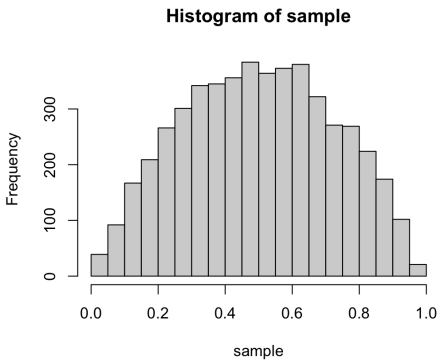
$$H(x, p) = -\log(x(1-x)) + \frac{p^2}{2}$$

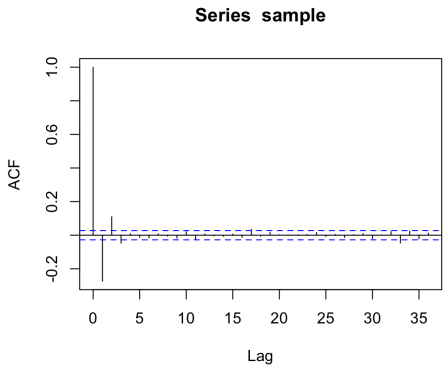
```
pdot = function(x){1/x-1/(1-x)}
h = function(x, p){
    if(x*(1-x) <= 0) return(Inf)
    else return(-log(x*(1-x)) + p**2/2)}

eps = 0.05
L = 10
n = 5000
burnin = 500
sample = rep(0, n+1)
x = 0.5
```

Example

```
for(i in 1:(n+burnin)){
    p = rnorm(1)
    h_old = h(x, p)
    y = x
    for(j in 1:L){
        p = p + 0.5 * pdot(y) * eps
        y = y + p * eps
        p = p + 0.5 * pdot(y) * eps
    }
    h_new = h(y, p)
    if(runif(1) <= exp(-h_new + h_old)) x = y
    if(i > burnin) sample[i-burnin] = x
}
```

Example


Histogram of sample
Frequency
0
100
200
300
0.0
0.2
0.4
0.6
0.8
1.0
sample


Series sample
ACF
-0.2
0.2
0.6
1.0
0
5
10
15
20
25
30
35
Lag