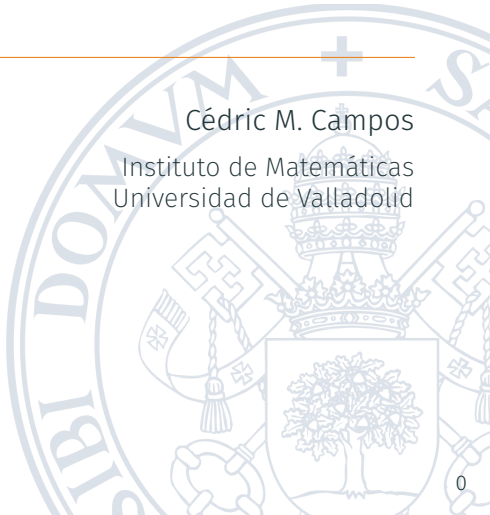
Extra Chance in Monte Carlo

Cédric M. Campos

Instituto de Matemáticas
Universidad de Valladolid

New Perspectives in Markov Chain Monte Carlo

8-12 June 2015, Valladolid (Spain)



Introduction & Motivation

Markov chain Monte Carlo methods

Family of methods for the simulation of stochastic processes.

How to: given Π (up to $\propto$), Metropolis-Hasting generates a Markov chain $\{X_n\}_{n\geq 0}$ for which Π is stationary.

Problem: correlation (the samples are not independent).

Solution: dynamics (gHMC) and, moreover, persistence (xHMC).

Markov chain Monte Carlo methods

Family of methods for the simulation of stochastic processes.

How to: given Π (up to $\propto$), Metropolis-Hasting generates a Markov chain $\{X_n\}_{n\geq 0}$ for which Π is stationary.

Problem: correlation (the samples are not independent).

Solution: dynamics (gHMC) and, moreover, persistence (xHMC).

Markov chain Monte Carlo methods

Family of methods for the simulation of stochastic processes.

How to: given Π (up to $\propto$), Metropolis-Hasting generates a Markov chain $\{X_n\}_{n\geq 0}$ for which Π is stationary.

Problem: correlation (the samples are not independent).

Solution: dynamics (gHMC) and, moreover, persistence (xHMC).

Markov chain Monte Carlo methods

Family of methods for the simulation of stochastic processes.

How to: given Π (up to $\propto$), Metropolis-Hasting generates a Markov chain $\{X_n\}_{n\geq 0}$ for which Π is stationary.

Problem: correlation (the samples are not independent).

Solution: dynamics (gHMC) and, moreover, persistence (xHMC).

Summary

- Given Π (up to $\propto$), we define the potential energy

$$V(q) = -\log(\Pi(q))$$

- We consider the (fictional) Hamiltonian

$$H(q,p) = \frac{1}{2}p^T M^{-1} p + V(q)$$

- Instead of sampling $\Pi(q) = \exp(-V(q))$ in state-space $q \in \mathbb{R}^d$, we sample

$$\Pi(q,p) = \exp(-H(q,p)) = \exp(-\frac{1}{2}p^T M^{-1} p) \cdot \exp(-V(q))$$

in phase-space $(q,p) \in \mathbb{R}^d \times \mathbb{R}^d$.

Note: q and p are independently distributed, $p \sim N(0, M)$.

- Dynamics described by

$$H(q,p) = \frac{1}{2}p^T M^{-1} p + V(q)$$

- Motions characterized by

$$\dot{q} = M^{-1}p\,, \quad \dot{p} = -\nabla V$$

- Can be hard to integrate

$$(q(\tau), p(\tau)) = \mathrm{L}_\tau(q_0, p_0)$$

  $\rightsquigarrow$ numerical integrators

- Challenge: conservation laws
  $\rightsquigarrow$ energy, symplecticity (volume)

Verlet / Leapfrog integrator

$$\begin{aligned} p_{1/2} &= p_0 = \frac{h}{2} f(q_0) \\ q_1 &= q_0 + h p_{1/2} \\ p_1 &= p_{1/2} + \frac{h}{2} f(q_1) \end{aligned}$$

where $f = -\nabla V$

$$\mathrm{L}_\tau \rightsquigarrow \mathrm{L}_{s,h}$$

gHMC & xHMC

```
01  p_0 = 0
02  ξ_0 = (q_0, p_0)
03  for n = 0 to N − 1 do
04    ξ_n^0 = R ξ_n
05    ξ_n^1 =   L ξ_n^0
06    Σ = min (1, Π(ξ_n^1)/Π(ξ_n^0))
07    if u_n ≤ Σ then
08      ξ_{n+1} =   ξ_n^1
09    else
10      ξ_{n+1} = F ξ_n^0
11    end_if
12    q_{n+1} = q(ξ_{n+1})
13  end_for
```

```
01  p_0 = 0
02  ξ_0 = (q_0, p_0)
03  for n = 0 to N − 1 do
04    ξ_n^0 = R ξ_n
05    ξ_n^1 =   L ξ_n^0
06    Σ = min (1, Π(ξ_n^1)/Π(ξ_n^0))
07    if u_n ≤ Σ then
08      ξ_{n+1} =   ξ_n^1
09    else
10      ξ_{n+1} = F ξ_n^0
11    end_if
12    q_{n+1} = q(ξ_{n+1})
13  end_for
```

Step 0: preparation

- Space extension

$$q \quad \Longrightarrow \quad \xi = (q, p)$$

- Hamiltonian dynamics

$$H(\xi) = \tfrac{1}{2} p^T M^{-1} p + V(q)$$

- Maxwell-Boltzmann distr.

$$\Pi(\xi) \propto \exp(-H(\xi))$$

- Start the loop

01 $p_0 = 0$
02 $\xi_0 = (q_0, p_0)$
03 **for** $n = 0$ **to** $N - 1$ **do**
04 $\quad \xi_n^0 = \mathrm{R}\xi_n$
05 $\quad \xi_n^1 = \mathrm{L}\xi_n^0$
06 $\quad \Sigma = \min\left(1, \Pi(\xi_n^1)/\Pi(\xi_n^0)\right)$
07 $\quad$ **if** $u_n \leq \Sigma$ **then**
08 $\quad\quad \xi_{n+1} = \xi_n^1$
09 $\quad$ **else**
10 $\quad\quad \xi_{n+1} = \mathrm{F}\xi_n^0$
11 $\quad$ **end_if**
12 $\quad q_{n+1} = q(\xi_{n+1})$
13 **end_for**

Step 0: preparation

- Space extension

$$q \quad \Longrightarrow \quad \xi = (q, p)$$

- Hamiltonian dynamics

$$H(\xi) = \tfrac{1}{2} p^T M^{-1} p + V(q)$$

- Maxwell-Boltzmann distr.

$$\Pi(\xi) \propto \exp(-H(\xi))$$

- Start the loop

```
01  p_0 = 0
02  ξ_0 = (q_0, p_0)
03  for n = 0 to N − 1 do
04    ξ_n^0 = Rξ_n
05    ξ_n^1 = Lξ_n^0
06    Σ = min(1, Π(ξ_n^1)/Π(ξ_n^0))
07    if u_n ≤ Σ then
08      ξ_{n+1} = ξ_n^1
09    else
10      ξ_{n+1} = Fξ_n^0
11    end_if
12    q_{n+1} = q(ξ_{n+1})
13  end_for
```

Step 1: momentum refreshment

$$\begin{aligned} \mathrm{R}\xi &= \mathrm{R}_\psi(q,p) := (q,p') \\ p' &:= \cos\psi \cdot p + \sin\psi \cdot z \\ z &\sim N(0,1) \end{aligned}$$

where $\psi \in [0, \pi/2] \rightsquigarrow \sin\psi \in [0,1]$

Note: if $\sin\psi = 1$ then $p' = p(\mathrm{R}\xi) \sim N(0,1)$

```
01  p_0 = 0
02  ξ_0 = (q_0, p_0)
03  for n = 0 to N − 1 do
04    ξ_n^0 = R ξ_n
05    ξ_n^1 = L ξ_n^0
06    Σ = min(1, Π(ξ_n^1)/Π(ξ_n^0))
07    if u_n ≤ Σ then
08      ξ_{n+1} = ξ_n^1
09    else
10      ξ_{n+1} = F ξ_n^0
11    end_if
12    q_{n+1} = q(ξ_{n+1})
13  end_for
```

Step 1: momentum refreshment

$$\begin{aligned} \mathrm{R}\xi &= \mathrm{R}_\psi(q,p) := (q,p') \\ p' &:= \cos\psi \cdot p + \sin\psi \cdot z \\ z &\sim N(0,1) \end{aligned}$$

where $\psi \in [0, \pi/2] \rightsquigarrow \sin\psi \in [0,1]$

Note: If $\sin\psi = 1$ then

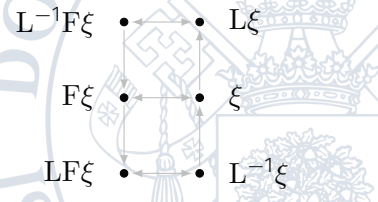
$$p' = p(\mathrm{R}\xi) \sim N(0,1)$$

01 $p_0 = 0$
02 $\xi_0 = (q_0, p_0)$
03 **for** $n = 0$ **to** $N - 1$ **do**
04 $\quad \xi_n^0 = \mathrm{R}\xi_n$
05 $\quad \xi_n^1 = \mathrm{L}\xi_n^0$
06 $\quad \Sigma = \min\left(1, \Pi(\xi_n^1)/\Pi(\xi_n^0)\right)$
07 $\quad$ **if** $u_n \leq \Sigma$ **then**
08 $\quad\quad \xi_{n+1} = \xi_n^1$
09 $\quad$ **else**
10 $\quad\quad \xi_{n+1} = \mathrm{F}\xi_n^0$
11 $\quad$ **end_if**
12 $\quad q_{n+1} = q(\xi_{n+1})$
13 **end_for**

Step 2: integrate & flip

$$\mathrm{L}\colon (q_0, p_0) \overset{\mathrm{L}_{s,h}}{\longmapsto} (q_1, p_1)$$

$$\mathrm{F}\colon (q, p) \longmapsto (q, -p)$$



01 $p_0 = 0$
02 $\xi_0 = (q_0, p_0)$
03 **for** $n = 0$ **to** $N - 1$ **do**
04 $\quad \xi_n^0 = \mathrm{R}\xi_n$
05 $\quad \xi_n^1 = \mathrm{L}\xi_n^0$
06 $\quad \Sigma = \min\left(1, \Pi(\xi_n^1)/\Pi(\xi_n^0)\right)$
07 $\quad$ **if** $u_n \leq \Sigma$ **then**
08 $\quad\quad \xi_{n+1} = \xi_n^1$
09 $\quad$ **else**
10 $\quad\quad \xi_{n+1} = \mathrm{F}\xi_n^0$
11 $\quad$ **end_if**
12 $\quad q_{n+1} = q(\xi_{n+1})$
13 **end_for**

Step 3: accept/reject

$$\xi_{n+1} = \begin{cases} \mathrm{L}\xi_n & \text{w. prob. } \Sigma \\ \mathrm{F}\xi_n & \text{w. prob. } 1 - \Sigma \end{cases}$$

by drawing $u_n \sim U(0, 1)$

01 $p_0 = 0$
02 $\xi_0 = (q_0, p_0)$
03 **for** $n = 0$ **to** $N - 1$ **do**
04 $\quad \xi_n^0 = \mathrm{R}\xi_n$
05 $\quad \xi_n^1 = \mathrm{L}\xi_n^0$
06 $\quad \Sigma = \min\left(1, \Pi(\xi_n^1)/\Pi(\xi_n^0)\right)$
07 $\quad$ **if** $u_n \leq \Sigma$ **then**
08 $\quad\quad \xi_{n+1} = \xi_n^1$
09 $\quad$ **else**
10 $\quad\quad \xi_{n+1} = \mathrm{F}\xi_n^0$
11 $\quad$ **end_if**
12 $\quad q_{n+1} = q(\xi_{n+1})$
13 **end_for**

Step 3: accept/reject

$$\xi_{n+1} = \begin{cases} \mathrm{L}\xi_n & \text{w. prob. } \Sigma \\ \mathrm{F}\xi_n & \text{w. prob. } 1 - \Sigma \end{cases}$$

by drawing $u_n \sim U([0, 1])$

$p_0 = 0$
$\xi_0 = (q_0, p_0)$
for $n = 0$ **to** $N - 1$ **do**
  $\xi_n^0 = \mathrm{R}\xi_n$
  $\xi_n^1 = \mathrm{L}\xi_n^0$
  $\Sigma = \min\left(1, \Pi(\xi_n^1)/\Pi(\xi_n^0)\right)$
  if $u_n \leq \Sigma$ **then**
    $\xi_{n+1} = \xi_n^1$
  else
    $\xi_{n+1} = \mathrm{F}\xi_n^0$
  end_if
  $q_{n+1} = q(\xi_{n+1})$
end_for

Step 4: throw momentum away

$p_0 = 0$
$\xi_0 = (q_0, p_0)$
for $n = 0$ **to** $N - 1$ **do**
  $\xi_n^0 = \mathrm{R}\xi_n$
  $\xi_n^1 = \mathrm{FL}\xi_n^0$
  $\Sigma = \min\left(1, \Pi(\xi_n^1)/\Pi(\xi_n^0)\right)$
  if $u_n \leq \Sigma$ **then**
    $\xi_{n+1} = \mathrm{F}\xi_n^1$
  else
    $\xi_{n+1} = \xi_n^0$
  end_if
  $q_{n+1} = q(\xi_{n+1})$
end_for

Finale: genHMC algorithm

Goal: improve acceptance rate (and chain quality).

Problem: rejections are costly.

Question: can we salvage them?

Goal: improve acceptance rate (and chain quality).

Problem: rejections are costly.

Question: can we salvage them?

Goal: improve acceptance rate (and chain quality).

Problem: rejections are costly.

Question: can we salvage them?

Goal: improve acceptance rate (and chain quality).

Problem: rejections are costly.

Question: can we salvage them?

Idea: generate an integration chain and choose (according to the dynamics)

$K \geq 1$ total chances

$K - 1 \geq 0$ extra chances

J. Sohl-Dickstein, M. Mudigonda & M.R. DeWeese
Hamiltonian Monte Carlo Without Detailed Balance.
Proc. Int. Conf. Mach. Learn. **31** (2014).

Goal: improve acceptance rate (and chain quality).

Problem: rejections are costly.

Question: can we salvage them?

$$\xi_{n+1} = \begin{cases} \mathrm{L}^1\xi_n \text{ with probability } \pi_{\mathrm{L}^1}(\xi_n) \\ \mathrm{L}^2\xi_n \text{ with probability } \pi_{\mathrm{L}^2}(\xi_n) \\ \qquad\vdots \\ \mathrm{L}^K\xi_n \text{ with probability } \pi_{\mathrm{L}^K}(\xi_n) \\ \mathrm{F}\xi_n \text{ with probability } \pi_{\mathrm{F}}(\xi_n) \end{cases}$$

J. Sohl-Dickstein, M. Mudigonda & M.R. DeWeese
Hamiltonian Monte Carlo Without Detailed Balance.
Proc. Int. Conf. Mach. Learn. **31** (2014).

$K \geq 1$ total chances
$K - 1 \geq 0$ extra chances

Definition (Cumulative Transition Probabilities [Campos & Sanz-Serna, 2015])

$$\begin{aligned} \Sigma^a(\xi) &:= \min\left\{1, \max_{b<a} \frac{\Pi(\mathrm{FL}^b\xi)}{\Pi(\xi)}\right\} \\ \pi_{\mathrm{L}^{a+1}}(\xi) &:= \Sigma^{a+1}(\xi) - \Sigma^a(\xi) \\ \pi_{\mathrm{F}}(\xi) &:= 1 - \Sigma^K \end{aligned}$$

Theorem (Detailed Balance [Campos & Sanz-Serna, 2015])

The associated transition kernel satisfies detailed balance.

Proof.

Thanks to reversibility and volume preservation, establish $\Pi(\xi)\cdot\pi_{\mathrm{L}^a}(\xi) =$
$= \Pi(\mathrm{FL}^a\xi)\cdot\pi_{\mathrm{L}^a}(\mathrm{FL}^a\xi)$. Then, go for the definition detailed balance. □

Definition (Cumulative Transition Probabilities [Campos & Sanz-Serna, 2015])

$$\begin{aligned} \Sigma^a(\xi) &:= \min\left\{1, \max_{b<a} \frac{\Pi(\mathrm{FL}^b\xi)}{\Pi(\xi)}\right\} \\ \pi_{\mathrm{L}^{a+1}}(\xi) &:= \Sigma^{a+1}(\xi) - \Sigma^a(\xi) \\ \pi_{\mathrm{F}}(\xi) &:= 1 - \Sigma^K \end{aligned}$$

Theorem (Detailed Balance [Campos & Sanz-Serna, 2015])

The associated transition kernel satisfies detailed balance.

Proof.

Thanks to reversibility and volume preservation, establish $\Pi(\xi)\cdot\pi_{\mathrm{L}^a}(\xi) =$
$= \Pi(\mathrm{FL}^a\xi)\cdot\pi_{\mathrm{L}^a}(\mathrm{FL}^a\xi)$. Then, go for the definition detailed balance. □

Definition (Cumulative Transition Probabilities [Campos & Sanz-Serna, 2015])

$$\begin{aligned} \Sigma^a(\xi) &:= \min\left\{1, \max_{b<a} \frac{\Pi(\mathrm{FL}^b\xi)}{\Pi(\xi)}\right\} \\ \pi_{\mathrm{L}^{a+1}}(\xi) &:= \Sigma^{a+1}(\xi) - \Sigma^a(\xi) \\ \pi_{\mathrm{F}}(\xi) &:= 1 - \Sigma^K \end{aligned}$$

Theorem (Detailed Balance [Campos & Sanz-Serna, 2015])

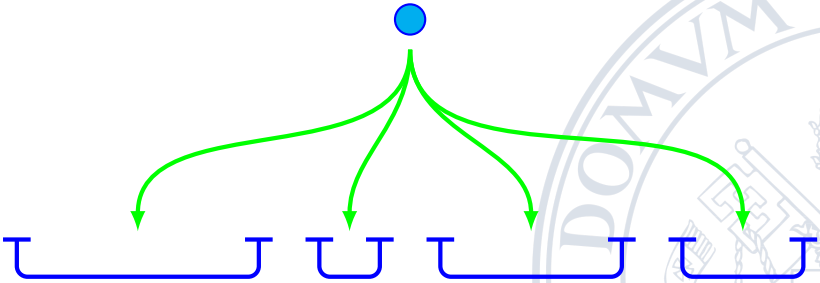
The associated transition kernel satisfies detailed balance.

Proof.

Thanks to reversibility and volume preservation, establish $\Pi(\xi) \cdot \pi_{\mathrm{L}^a}(\xi) =$ $= \Pi(\mathrm{FL}^a\xi) \cdot \pi_{\mathrm{L}^a}(\mathrm{FL}^a\xi)$. Then, go for the definition detailed balance. □

$$\Sigma^a(\xi) = \max_{b<a} \min \left\{ 1,\ \frac{\Pi(\mathrm{FL}^b \xi)}{\Pi(\xi)} \right\} = \sum_a \pi_{\mathrm{L}^a}(\xi)$$

$$\Sigma^a(\xi) = \max_{b<a} \min \left\{ 1,\ \frac{\Pi(\mathrm{FL}^b \xi)}{\Pi(\xi)} \right\} = \sum_a \pi_{\mathrm{L}^a}(\xi)$$

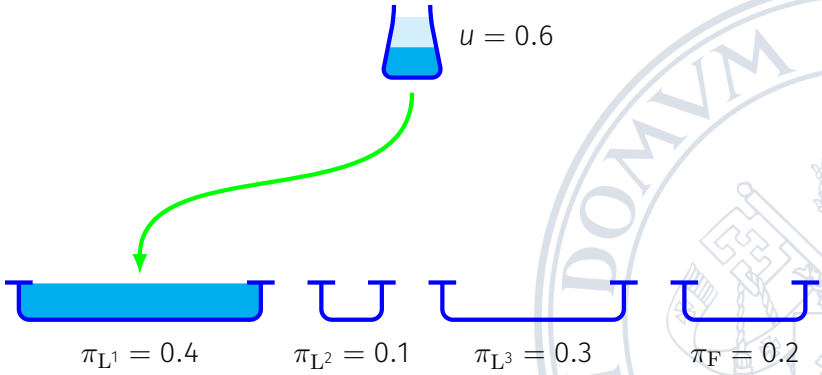


$$\Sigma^a(\xi) = \max_{b<a} \min \left\{ 1,\ \frac{\Pi(\mathrm{FL}^b \xi)}{\Pi(\xi)} \right\} = \sum_a \pi_{\mathrm{L}^a}(\xi)$$

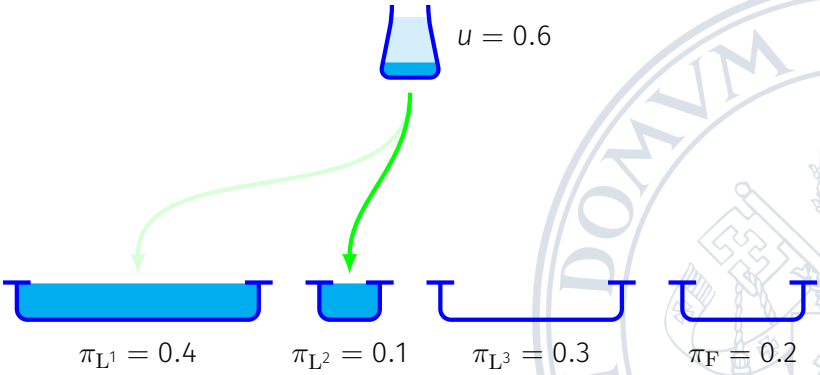
$u = 0.6$

$\pi_{\mathrm{L}^1} = 0.4$ $\pi_{\mathrm{L}^2} = 0.1$ $\pi_{\mathrm{L}^3} = 0.3$ $\pi_{\mathrm{F}} = 0.2$

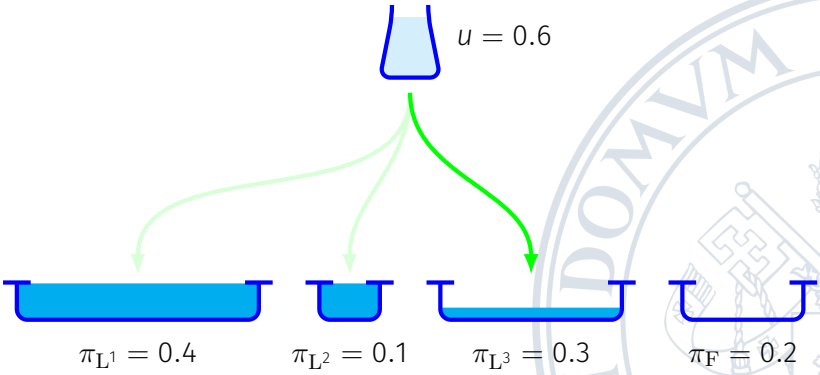
$$\Sigma^a(\xi) = \max_{b<a} \min \left\{ 1,\ \frac{\Pi(\mathrm{FL}^b \xi)}{\Pi(\xi)} \right\} = \sum_a \pi_{\mathrm{L}^a}(\xi)$$



$$\Sigma^a(\xi) = \max_{b<a} \min \left\{ 1,\ \frac{\Pi(\mathrm{FL}^b \xi)}{\Pi(\xi)} \right\} = \sum_a \pi_{\mathrm{L}^a}(\xi)$$



$$\Sigma^a(\xi) = \max_{b<a} \min \left\{ 1,\ \frac{\Pi(\mathrm{FL}^b \xi)}{\Pi(\xi)} \right\} = \sum_a \pi_{\mathrm{L}^a}(\xi)$$



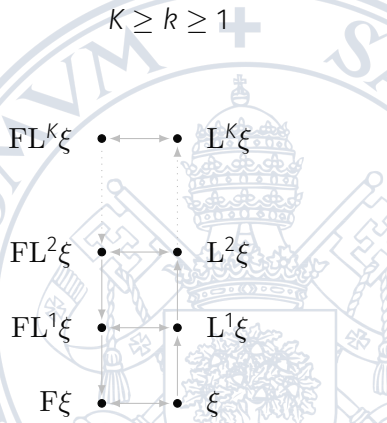
```
01  p_0 = 0
02  ξ_0 = (q_0, p_0)
03  for n = 0 to N − 1 do
04    ξ_n^0 = R ξ_n
05    Σ = k = 0
06    do
07      k = k + 1
08      ξ_n^k = L ξ_n^{k−1}
09      Σ = max{Σ, min{1, Π(ξ_n^k)/Π(ξ_n^0)}}
10    while u_n > Σ and k < K
11    if u_n ≤ Σ then
12      ξ_{n+1} = ξ_n^k
13    else
14      ξ_{n+1} = F ξ_n^0
15    end_if
16    q_{n+1} = q(ξ_{n+1})
17  end_for
```

genHMC: succeed or resign

$K = k = 1$

```
01  p_0 = 0
02  ξ_0 = (q_0, p_0)
03  for n = 0 to N − 1 do
04    ξ_n^0 = R ξ_n
05    Σ = k = 0
06    do
07      k = k + 1
08      ξ_n^k = L ξ_n^{k−1}
09      Σ = max{Σ, min{1, Π(ξ_n^k)/Π(ξ_n^0)}}
10    while u_n > Σ and k < K
11    if u_n ≤ Σ then
12      ξ_{n+1} = ξ_n^k
13    else
14      ξ_{n+1} = F ξ_n^0
15    end_if
16    q_{n+1} = q(ξ_{n+1})
17  end_for
```

xHMC: persist



Quantity & Quality

Effective Sample Size (*ess*)

How to quantitatively determine the meaningful data of an MC?

$X_1, X_2, \ldots, X_m$ MC $\sim \Pi$

$$\mu_m^X \to \mu$$

$$\sqrt{m}(\mu_m^X - \mu) \to N(0, \sigma^2)$$

$$\mu = \int g(x)\Pi(x)\,\mathrm{d}x$$

$$\mu_r^Z = \frac{1}{r}\sum_{k=1}^{r} g(Z_k)$$

$Y_1, Y_2, \ldots, Y_n$ i.i.d. $\sim \Pi$

$$\mu_n^Y \to \mu$$

$$\sqrt{n}(\mu_n^Y - \mu) \to N(0, s^2)$$

Effective Sample Size (*ess*)

How to quantitatively determine the meaningful data of an MC?

$X_1, X_2, \ldots, X_m$ MC $\sim \Pi$

$$\mu_m^X \to \mu$$

$$\sqrt{m}(\mu_m^X - \mu) \to N(0, \sigma^2)$$

$$\mu = \int g(x)\Pi(x)\,\mathrm{d}x$$

$$\mu_r^Z = \frac{1}{r}\sum_{k=1}^{r} g(Z_k)$$

$Y_1, Y_2, \ldots, Y_n$ i.i.d. $\sim \Pi$

$$\mu_n^Y \to \mu$$

$$\sqrt{n}(\mu_n^Y - \mu) \to N(0, s^2)$$

How to quantitatively determine the meaningful data of an MC?

$X_1, X_2, \ldots, X_m$ MC $\sim \Pi$

$\mu_m^X \to \mu$

$\sqrt{m}(\mu_m^X - \mu) \to N(0, \sigma^2)$

$\mu = \int g(x)\Pi(x)\,\mathrm{d}x$

$\mu_r^Z = \frac{1}{r}\sum_{k=1}^{r} g(Z_k)$

$Y_1, Y_2, \ldots, Y_n$ i.i.d. $\sim \Pi$

$\mu_n^Y \to \mu$

$\sqrt{n}(\mu_n^Y - \mu) \to N(0, s^2)$

How to quantitatively determine the meaningful data of an MC?

$X_1, X_2, \ldots, X_m$ MC $\sim \Pi$

$$\mu_m^X \to \mu$$

$$\sqrt{m}(\mu_m^X - \mu) \to N(0, \sigma^2)$$

$$\mu = \int g(x)\Pi(x)\,\mathrm{d}x$$

$$\mu_r^Z = \frac{1}{r}\sum_{k=1}^{r} g(Z_k)$$

$Y_1, Y_2, \ldots, Y_n$ i.i.d. $\sim \Pi$

$$\mu_n^Y \to \mu$$

$$\sqrt{n}(\mu_n^Y - \mu) \to N(0, s^2)$$

How to quantitatively determine the meaningful data of an MC?

$X_1, X_2, \ldots, X_m$ MC $\sim \Pi$

$$\mu_m^X \to \mu$$

$$\sqrt{m}(\mu_m^X - \mu) \to N(0, \sigma^2)$$

$$\mu = \int g(x)\Pi(x)\,\mathrm{d}x$$

$$\mu_r^Z = \frac{1}{r}\sum_{k=1}^{r} g(Z_k)$$

$Y_1, Y_2, \ldots, Y_n$ i.i.d. $\sim \Pi$

$$\mu_n^Y \to \mu$$

$$\sqrt{n}(\mu_n^Y - \mu) \to N(0, s^2)$$

How to quantitatively determine the meaningful data of an MC?

$X_1, X_2, \ldots, X_m$ MC $\sim \Pi$

$\mu_m^X \to \mu$

$\sqrt{m}(\mu_m^X - \mu) \to N(0, \sigma^2)$

$\mu = \int g(x)\Pi(x)\,\mathrm{d}x$

$\mu_r^Z = \frac{1}{r}\sum_{k=1}^{r} g(Z_k)$

$Y_1, Y_2, \ldots, Y_n$ i.i.d. $\sim \Pi$

$\mu_n^Y \to \mu$

$\sqrt{n}(\mu_n^Y - \mu) \to N(0, s^2)$

$$\frac{\sigma^2}{m} \approx \frac{s^2}{n} \qquad \Longrightarrow \qquad n \approx m\frac{s^2}{\sigma^2}$$

$\sigma^2 \approx ?¿$

$$s^2 \approx \frac{1}{m}\sum_{i=1}^{0}(g(X_i) - \mu_m^X)^2 =: \gamma_{m,0}^X$$

How to quantitatively determine the meaningful data of an MC?

$X_1, X_2, \ldots, X_m$ MC $\sim \Pi$

$\mu_m^X \to \mu$

$\sqrt{m}(\mu_m^X - \mu) \to N(0, \sigma^2)$

$\mu = \int g(x)\Pi(x)\,\mathrm{d}x$

$\mu_r^Z = \frac{1}{r}\sum_{k=1}^{r} g(Z_k)$

$Y_1, Y_2, \ldots, Y_n$ i.i.d. $\sim \Pi$

$\mu_n^Y \to \mu$

$\sqrt{n}(\mu_n^Y - \mu) \to N(0, s^2)$

Theorem (Kipnis & Varadahan, 1986)

For a stationary, irreducible, reversible Markov chain

$$m \operatorname{var} \mu_m^X \to \sigma^2 := \sum_{-\infty}^{+\infty} \gamma_k\,, \quad \gamma_k := \operatorname{cov}(g(X_i), g(X_{i+k}))\,.$$

If σ^2 is finite, then

$$\sqrt{m}(\mu_m^X - \mu) \to N(0, \sigma^2)$$

Effective Sample Size (*ess*)

- Direct *approaches* to compute γ_k give bad results.
- Alternative, define $\Gamma_k := \gamma_{2k} + \gamma_{2k+1}$

Theorem (Geyer, 1992)

$\Gamma_k = \Gamma(k)$ is strictly positive, strictly decreasing and strictly convex.

- Approximate $\gamma_k \approx \gamma^X_{m,k} := \frac{1}{m}\sum_{i=1}^{m-k}(g(X_i) - \mu^X_m)(g(X_{i+k}) - \mu^X_m)$

  and $\Gamma_k \approx \Gamma^X_{m,k} := \gamma^X_{m,2k} + \gamma^X_{m,2k+1}$
- Modify $\Gamma^X_{m,k}$ such that theorem is "satisfied".
- Then $\sigma^2 \approx -\gamma^X_{m,0} + 2\sum \Gamma^X_{m,k}$.

C.J. Geyer
Practical Markov Chain Monte Carlo.
Statistical Science **7** (1992).

- Direct *approaches* to compute γ_k give bad results.
- Alternative, define $\Gamma_k := \gamma_{2k} + \gamma_{2k+1}$

Theorem (Geyer, 1992)

$\Gamma_k = \Gamma(k)$ *is strictly positive, strictly decreasing and strictly convex.*

- Approximate $\gamma_k \approx \gamma^X_{m,k} := \frac{1}{m}\sum_{i=1}^{m-k}(g(X_i) - \mu^X_m)(g(X_{i+k}) - \mu^X_m)$

  and $\Gamma_k \approx \Gamma^X_{m,k} := \gamma^X_{m,2k} + \gamma^X_{m,2k+1}$
- Modify $\Gamma^X_{m,k}$ such that theorem is "satisfied".
- Then $\sigma^2 \approx -\gamma^X_{m,0} + 2\sum \Gamma^X_{m,k}$.

C.J. Geyer
Practical Markov Chain Monte Carlo.
Statistical Science **7** (1992).

- Direct *approaches* to compute γ_k give bad results.
- Alternative, define $\Gamma_k := \gamma_{2k} + \gamma_{2k+1}$

Theorem (Geyer, 1992)

$\Gamma_k = \Gamma(k)$ *is strictly positive, strictly decreasing and strictly convex.*

- Approximate $\gamma_k \approx \gamma^X_{m,k} := \frac{1}{m}\sum_{i=1}^{m-k}(g(X_i) - \mu^X_m)(g(X_{i+k}) - \mu^X_m)$

  and $\Gamma_k \approx \Gamma^X_{m,k} := \gamma^X_{m,2k} + \gamma^X_{m,2k+1}$
- Modify $\Gamma^X_{m,k}$ such that theorem is "satisfied".
- Then $\sigma^2 \approx -\gamma^X_{m,0} + 2\sum \Gamma^X_{m,k}$.

C.J. Geyer
Practical Markov Chain Monte Carlo.
Statistical Science **7** (1992).

Case of Study & Benchmark

Linear alkanes are chemical compounds of the form

$$CH_3 - CH_2 - \cdots - CH_2 - CH_3$$

The conformation of the molecule is characterized by the positions of the Carbon atoms.

E. Cancès, F. Legoll & G. Stoltz
Theoretical and numerical comparison of some sampling methods for molecular dynamics.
ESAIM: M2AN **41** (2007).

It is assumed they behave according to the potential

$$V(q) = \sum_{i=1}^{n-1} V_2(d_{i,i+1}) + \sum_{i=1}^{n-2} V_3(\theta_i) + \sum_{i=1}^{n-3} V_4(\phi_i) + \sum_{i=1}^{n-4} \sum_{j=i+3}^{n} V_{LJ}(d_{i,j})$$

Linear alkanes are chemical compounds of the form

$$CH_3 - CH_2 - CH_2 - CH_2 - \cdots - CH_2 - CH_3$$

It is assumed they behave according to the potential

$$V(q) = \sum_{i=1}^{n-1} V_2(d_{i,i+1}) + \sum_{i=1}^{n-2} V_3(\theta_i) + \sum_{i=1}^{n-3} V_4(\phi_i) + \sum_{i=1}^{n-4} \sum_{j=i+3}^{n} V_{LJ}(d_{i,j})$$

Linear alkanes are chemical compounds of the form

$$CH_3 - CH_2 - CH_2 - CH_2 - \cdots - CH_2 - CH_3$$

It is assumed they behave according to the potential

$$V(q) = \sum_{i=1}^{n-1} V_2(d_{i,i+1}) + \sum_{i=1}^{n-2} V_3(\theta_i) + \sum_{i=1}^{n-3} V_4(\phi_i) + \sum_{i=1}^{n-4} \sum_{j=i+3}^{n} V_{LJ}(d_{i,j})$$

where

Linear alkanes are chemical compounds of the form

$$CH_3 - CH_2 - CH_2 - CH_2 - \cdots - CH_2 - CH_3$$

It is assumed they behave according to the potential

$$V(q) = \sum_{i=1}^{n-1} V_2(d_{i,i+1}) + \sum_{i=1}^{n-2} V_3(\theta_i) + \sum_{i=1}^{n-3} V_4(\phi_i) + \sum_{i=1}^{n-4} \sum_{j=i+3}^{n} V_{LJ}(d_{i,j})$$

where

$$V_2(d) = \frac{1}{2} k_0 (d - d_0)^2$$

covalent bond potential between two consecutive Carbon atoms

- d, distance between the atoms
- k_0, d_0, system parameters

Linear alkanes are chemical compounds of the form

$$CH_3 - CH_2 - CH_2 - CH_2 - \cdots - CH_2 - CH_3$$

It is assumed they behave according to the potential

$$V(q) = \sum_{i=1}^{n-1} V_2(d_{i,i+1}) + \sum_{i=1}^{n-2} V_3(\theta_i) + \sum_{i=1}^{n-3} V_4(\phi_i) + \sum_{i=1}^{n-4} \sum_{j=i+3}^{n} V_{LJ}(d_{i,j})$$

where

$$V_3(\theta) = \frac{1}{2}\kappa_0(\theta - \theta_0)^2$$

three-body interaction potential

- θ, bending angle between three consecutive atoms
- κ_0, θ_0, system parameters

Linear alkanes are chemical compounds of the form

$$CH_3 - CH_2 - CH_2 - CH_2 - \cdots - CH_2 - CH_3$$

It is assumed they behave according to the potential

$$V(q) = \sum_{i=1}^{n-1} V_2(d_{i,i+1}) + \sum_{i=1}^{n-2} V_3(\theta_i) + \sum_{i=1}^{n-3} V_4(\phi_i) + \sum_{i=1}^{n-4} \sum_{j=i+3}^{n} V_{LJ}(d_{i,j})$$

where

$$V_4(\phi) = c_1(1-x) + 2c_2(1-x^2) + c_3(1+3x-4x^3)$$

four-body interaction potential

- $\phi = \arccos(x)$, diedral angle between four consecutive atoms
- c_1, c_2, c_3, system parameters

Linear alkanes are chemical compounds of the form

$$CH_3 - CH_2 - CH_2 - CH_2 - \cdots - CH_2 - CH_3$$

It is assumed they behave according to the potential

$$V(q) = \sum_{i=1}^{n-1} V_2(d_{i,i+1}) + \sum_{i=1}^{n-2} V_3(\theta_i) + \sum_{i=1}^{n-3} V_4(\phi_i) + \sum_{i=1}^{n-4} \sum_{j=i+3}^{n} V_{LJ}(d_{i,j})$$

where

$$V_{LJ}(d) = 4\epsilon \left(\left(\frac{\sigma}{d} \right)^{12} - \left(\frac{\sigma}{d} \right)^{6} \right)$$

Lennard-Jones potential of two Carbon atoms connected by three or more covalent bonds

- d, distance between the atoms
- ϵ, σ, system parameters

- alkane $= 9$ carbons (i.e. 6 diedra)
- $K = 0,\ 3$ (extra chances)
- $\sin\psi = 0.10,\ 0.25,\ 0.50,\ 0.75,\ 1$
- $L\Delta t = 0.12,\ 0.24,\ 0.36,\ 0.48,\ 0.60,\ 0.72$
- $\Delta t = 0.012,\ 0.020,\ 0.024,\ 0.030$
- chains per setting $= 10$

total chains $= 1800$ (several GiB)

- cost = 10^6 force evaluations
- ESS measures:
  - 1st diedra, $g(x) = \mathbf{1}_{[-1.75,1.75]}(x)$

```
92.6707 9240    3368    3632
77.0985 8853    3782    3965
65.3370 7431    3401    3518
92.9657 5.8980  0.9573  0.1426
80.2533 15.9591 3.2026  0.4637
71.6308 21.9310 5.3675  0.8597
92.5993 8943    3436    3504
77.1728 8822    4160    4096
65.2272 8421    4352    4146
92.8594 6.0066  0.9512  0.1486
80.0717 16.0101 3.3024  0.4854
71.4885 22.0166 5.4070  0.8759
92.5784 9952    4689    4510
77.2683 10530   5278    5117
65.3837 9445    4997    4745
92.9053 5.9769  0.9382  0.1476
79.9234 16.1746 3.2781  0.4934
71.2970 22.0998 5.5346  0.8526
92.5878 8169    4135    3961
77.2775 9736    5410    5231
65.5464 8982    5168    4937
92.8834 5.9911  0.9473  0.1382
```

Δt	$K=0$	$K=3$				
		a_0	a_1	a_2	a_3	$\sum a_k$
0.012	93%	93%	6%	1%	0%	100%
0.016	86%	87%	11%	2%	0%	100%
0.020	77%	80%	16%	3%	1%	100%
0.024	65%	71%	22%	6%	1%	100%

Acceptance rates (rounded) for $\sin\psi = 1$, $L\Delta t = 0.48$ and different values of Δt and K, $K = 0$ (standard HMC) and $K = 3$ (three extra chances). The column labelled a_k gives the acceptance rate at the $(k+1)$-try.

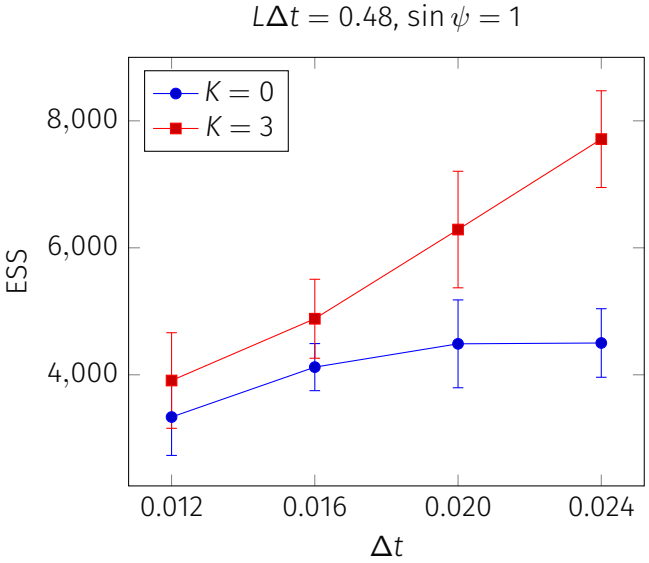
Conclusion: extra chance increases acceptance ratio

Question: worth? (tricky, shorter chain lengths)

Δt	$K=0$	$K=3$				
		a_0	a_1	a_2	a_3	$\sum a_k$
0.012	93%	93%	6%	1%	0%	100%
0.016	86%	87%	11%	2%	0%	100%
0.020	77%	80%	16%	3%	1%	100%
0.024	65%	71%	22%	6%	1%	100%

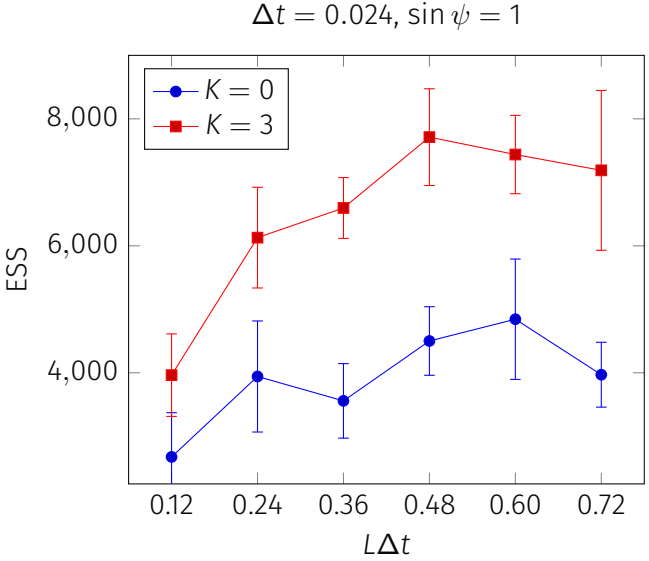
Acceptance rates (rounded) for $\sin\psi = 1$, $L\Delta t = 0.48$ and different values of Δt and K, $K = 0$ (standard HMC) and $K = 3$ (three extra chances). The column labelled a_k gives the acceptance rate at the $(k+1)$-try.

Conclusion: extra chance increases acceptance ratio.
Question: worth? (tricky, shorter chain lengths)



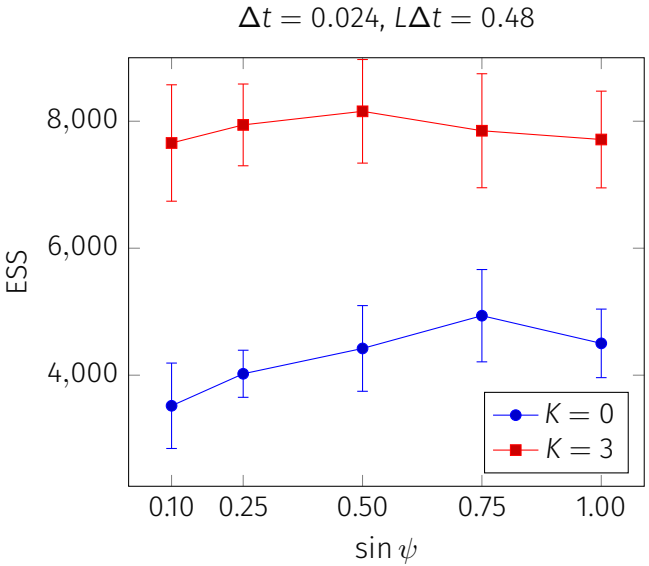
ESS as a function of ITS

- fixed ITI
- full mom. refresh



ESS as a function of ITI

- fixed ITS
- full momentum refreshment



ESS as a function of refreshment

- fixed ITS
- fixed ITI

Epi & Logue

Conclusions:

- clever (uses dynamics)
- applicable (detailed balance)
- easily implementable (simple back-recurrence)
- worth (quality chains)

Outlook:

- integration (cheap & tailored)

Extra:

- Matlab/Octave code available at http://github.com/vitaminace33

C.M. Campos & J.M. Sanz-Serna
Extra Chance Generalized Hybrid Monte Carlo.
J. Comp. Phys. **281** (2015).

E. Cancès, F. Legoll & G. Stoltz
Theoretical and numerical comparison of some sampling methods for molecular dynamics.
ESAIM: M2AN **41** (2007).

C.J. Geyer
Practical Markov Chain Monte Carlo.
Statistical Science **7** (1992).

J. Sohl-Dickstein, M. Mudigonda & M.R. DeWeese
Hamiltonian Monte Carlo Without Detailed Balance.
Proc. Int. Conf. Mach. Learn. **31** (2014).

Questions?

Copyleft, All Rights Reversed