Metropolis–Hastings
Numerical examples

Computational Statistics

The Metropolis–Hastings update

Let $\pi(x) = \tilde{\pi}(x)/Z$ be the target density. Given the current state $X_t = x$:

1. Propose $Y \sim q(\cdot \mid x)$.
2. Calculate
$$\alpha(x, Y) = \min\left\{1, \frac{\tilde{\pi}(Y)q(x \mid Y)}{\tilde{\pi}(x)q(Y \mid x)}\right\}.$$
3. Independently draw $U \sim \mathrm{Unif}(0, 1)$ and set
$$X_{t+1} = \begin{cases} Y, & U < \alpha(x, Y), \\ x, & U \geq \alpha(x, Y). \end{cases}$$

The unknown normalising constant Z cancels in the ratio.

Example 1: a mixture of three Gaussians

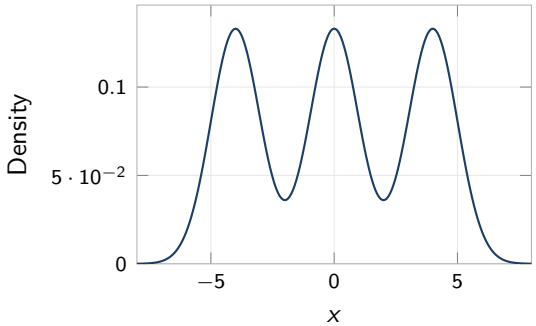
An equally weighted mixture of three Gaussians, each with variance 1:

$$\pi(x) = \tfrac{1}{3}\big[\varphi(x+4) + \varphi(x) + \varphi(x-4)\big].$$

Here φ is the standard normal density.

$$\mathbb{E}[X] = 0, \qquad \mathbb{E}[X^2] = \frac{35}{3}.$$

Here $\mathbb{E}[X^2] = 1 + (16 + 0 + 16)/3$. The target has support on all of $\mathbb{R}$.



Centres: $-4, 0, 4$. Each component has weight $1/3$.

A uniform random-walk proposal

Given $X_t = x$, draw two independent uniforms:

$$V \sim \mathrm{Unif}(-1, 1), \qquad U \sim \mathrm{Unif}(0, 1).$$

Propose $Y = x + hV$, so $Y \mid x \sim \mathrm{Unif}(x - h, x + h)$.

$$q(y \mid x) = \frac{1}{2h} \mathbf{1}\{|y - x| \le h\} = q(x \mid y).$$

For every proposed pair, the proposal densities cancel:

$$\boxed{\alpha(x, y) = \min\{1, \pi(y)/\pi(x)\}.}$$

Here h is the **half-width**, not the standard deviation $(h/\sqrt{3})$. The interval moves with the chain, so no fixed truncation of the target is imposed.

One acceptance decision, step by step

Use $h = 5$. Suppose $x = -4$, propose $y = -2$, and draw $U = 0.40$.

$$\pi(-4) \approx 0.133025, \qquad \pi(-2) \approx 0.035994.$$

The forward and reverse proposal densities both equal $1/10$, hence

$$r = \frac{\pi(-2)}{\pi(-4)} \approx 0.2706, \qquad \alpha(-4, -2) = 0.2706.$$

Since $0.40 > 0.2706$, reject and record $X_{t+1} = -4$ again.

If instead $U = 0.10$, accept and record $X_{t+1} = -2$. A proposal $-4 \to 0$ has ratio $\pi(0)/\pi(-4) \approx 1.0003$ and is always accepted.

A complete short chain

Start at $X_0 = -4$ and use $h = 5$. All proposed moves lie in the current interval.

t	X_t	Y	$\alpha(X_t, Y)$	U	X_{t+1}
0	-4	-2	0.2706	0.40	-4 (reject)
1	-4	0	1.0000	0.30	0 (accept)
2	0	2	0.2705	0.90	0 (reject)
3	0	4	0.9997	0.20	4 (accept)
4	4	6	0.1353	0.70	4 (reject)
5	4	3	0.6174	0.60	3 (accept)

These illustrative proposals and uniforms give the chain $-4, 0, 0, 4, 4, 3$. The acceptance rate is $3/6$, while the chain contains six recorded states.

Rejections contribute to Monte Carlo estimates

For a quantity of interest $\mathcal{I} = \mathbb{E}_\pi[f(X)]$, use

$$\widehat{\mathcal{I}}_N = \frac{1}{N}\sum_{t=1}^{N} f(X_t).$$

For the six-state chain and $f(x) = x$:

$$\widehat{\mathcal{I}}_6 = \frac{-4+0+0+4+4+3}{6} = \frac{7}{6} \approx 1.1667.$$

Repeated values encode time spent at a state

Discarding rejected updates changes the empirical distribution. In general, the accepted-move chain alone does not have stationary distribution π.

Six updates cannot establish convergence or give a reliable estimate. By symmetry, the exact mean of this mixture is $\mathcal{I} = 0$.

An evolving example: one chain from its first update

Target: $\pi(x) = \{\varphi(x+4) + \varphi(x) + \varphi(x-4)\}/3$.

$$X_0 = -4, \qquad Y = X_t + 5V, \quad V \sim \text{Unif}(-1, 1).$$

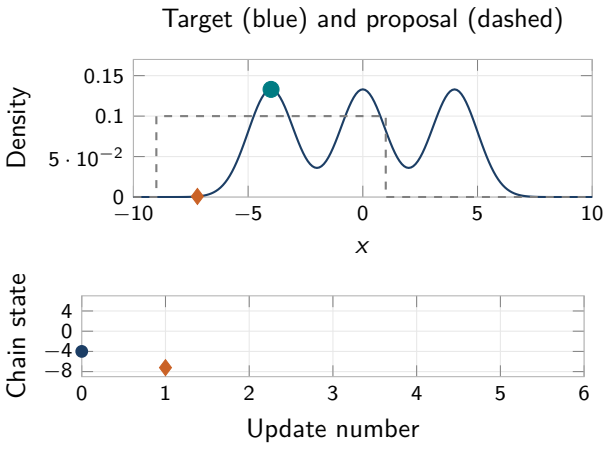
- For each of the first six updates, the first view shows the proposal.
- Advancing once reveals the acceptance calculation, uniform draw and next state.
- The trace grows after every decision, including a repeated state after rejection.

We then continue *the same run* to 5,000 updates and watch its histogram evolve. No initial updates are discarded. Decisions use full precision.
Teal dots mark current states, and orange diamonds mark proposals.

An evolving example: propose, decide, record

Update 1 of 6: Proposal

$X_0 = -4.0000$



Uniform half-width $h = 5$. Candidate $Y = -7.2107$.

The orange diamond is a candidate. It has not yet entered the chain.

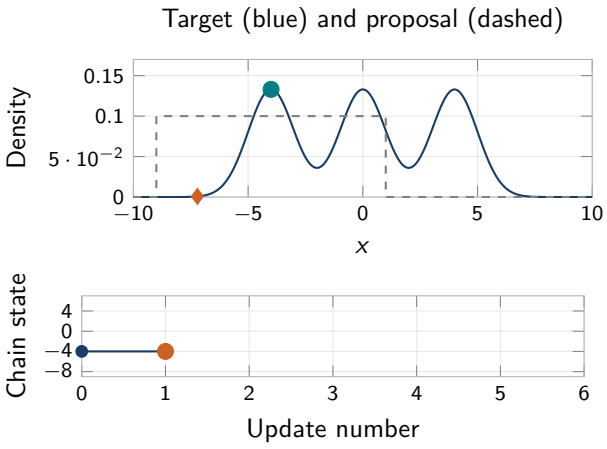
$$\alpha(X_t, Y) = \min\{1, \pi(Y)/\pi(X_t)\}.$$

Will the chain move, or repeat its current value?

Advance once to reveal the decision.

An evolving example: propose, decide, record

Update 1 of 6: Decision



$X_0 = -4.0000$

Uniform half-width $h = 5$. Candidate $Y = -7.2107$.

$$\log r = \log \frac{\pi(Y)}{\pi(X_t)} = -5.1545.$$

$$\alpha = \min\{1, e^{-5.1545}\} = 0.0058.$$

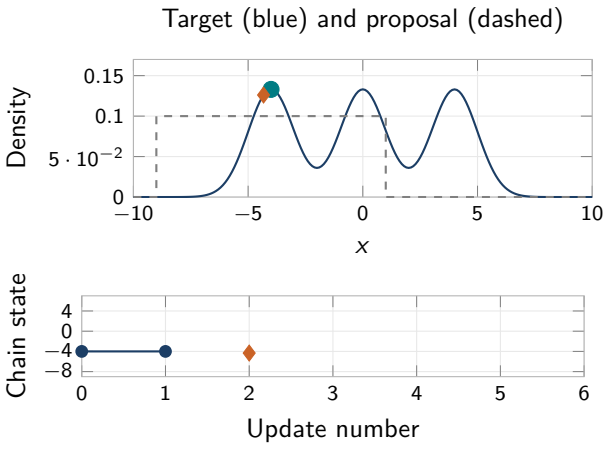
$$U = 0.6399 \geq \alpha.$$

Reject: $X_1 = -4.0000$.
Append the current value again. The repeated point is a sample.

An evolving example: propose, decide, record

Update 2 of 6: Proposal

$X_1 = -4.0000$



Uniform half-width $h = 5$. Candidate $Y = -4.3273$.

The orange diamond is a candidate. It has not yet entered the chain.

$$\alpha(X_t, Y) = \min\{1, \pi(Y)/\pi(X_t)\}.$$

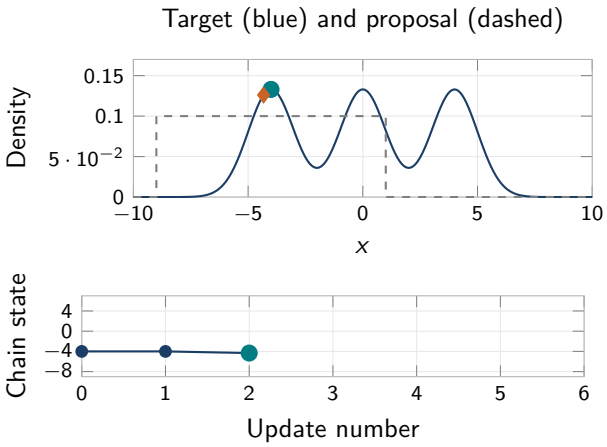
Will the chain move, or repeat its current value?

Advance once to reveal the decision.

An evolving example: propose, decide, record

Update 2 of 6: Decision

$X_1 = -4.0000$



Uniform half-width $h = 5$. Candidate $Y = -4.3273$.

$$\log r = \log \frac{\pi(Y)}{\pi(X_t)} = -0.0538.$$

$$\alpha = \min\{1, e^{-0.0538}\} = 0.9476.$$

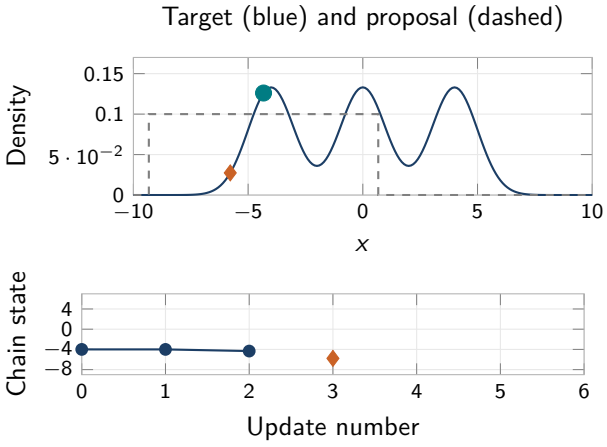
$$U = 0.3705 < \alpha.$$

Accept: $X_2 = -4.3273$.
Append the new value to the chain.

An evolving example: propose, decide, record

Update 3 of 6: Proposal

$X_2 = -4.3273$



Uniform half-width $h = 5$. Candidate $Y = -5.7781$.

The orange diamond is a candidate. It has not yet entered the chain.

$$\alpha(X_t, Y) = \min\{1, \pi(Y)/\pi(X_t)\}.$$

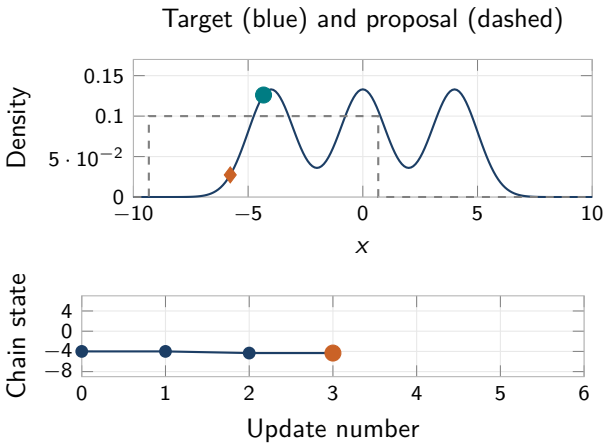
Will the chain move, or repeat its current value?

Advance once to reveal the decision.

An evolving example: propose, decide, record

Update 3 of 6: Decision

$X_2 = -4.3273$



Uniform half-width $h = 5$. Candidate $Y = -5.7781$.

$$\log r = \log \frac{\pi(Y)}{\pi(X_t)} = -1.5274.$$

$$\alpha = \min\{1, e^{-1.5274}\} = 0.2171.$$

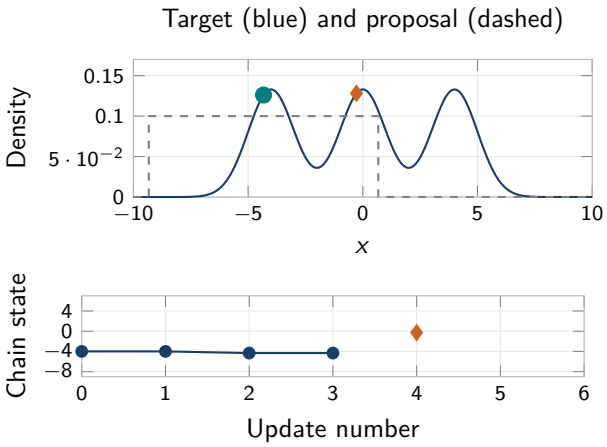
$$U = 0.7905 \geq \alpha.$$

Reject: $X_3 = -4.3273$.
Append the current value again. The repeated point is a sample.

An evolving example: propose, decide, record

Update 4 of 6: Proposal

$X_3 = -4.3273$



Uniform half-width $h = 5$. Candidate $Y = -0.2759$.

The orange diamond is a candidate. It has not yet entered the chain.

$$\alpha(X_t, Y) = \min\{1, \pi(Y)/\pi(X_t)\}.$$

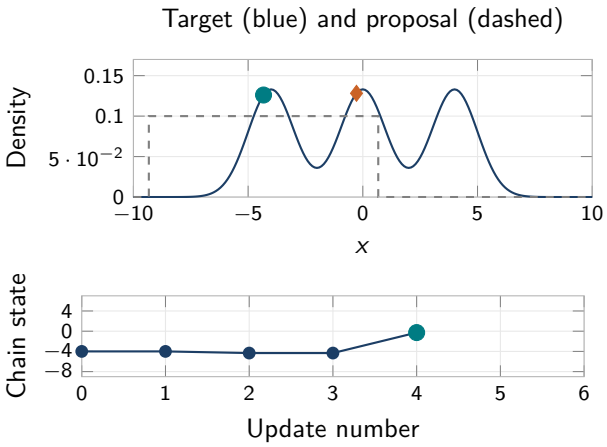
Will the chain move, or repeat its current value?

Advance once to reveal the decision.

An evolving example: propose, decide, record

Update 4 of 6: Decision

$X_3 = -4.3273$



Uniform half-width $h = 5$. Candidate $Y = -0.2759$.

$$\log r = \log \frac{\pi(Y)}{\pi(X_t)} = 0.0165.$$

$$\alpha = \min\{1, e^{0.0165}\} = 1.0000.$$

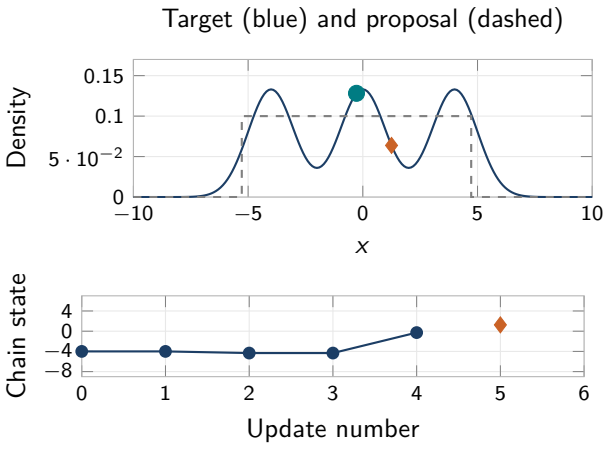
$$U = 0.1774 < \alpha.$$

Accept: $X_4 = -0.2759$.
Append the new value to the chain.

An evolving example: propose, decide, record

Update 5 of 6: Proposal

$X_4 = -0.2759$



Uniform half-width $h = 5$. Candidate $Y = 1.2520$.

The orange diamond is a candidate. It has not yet entered the chain.

$$\alpha(X_t, Y) = \min\{1, \pi(Y)/\pi(X_t)\}.$$

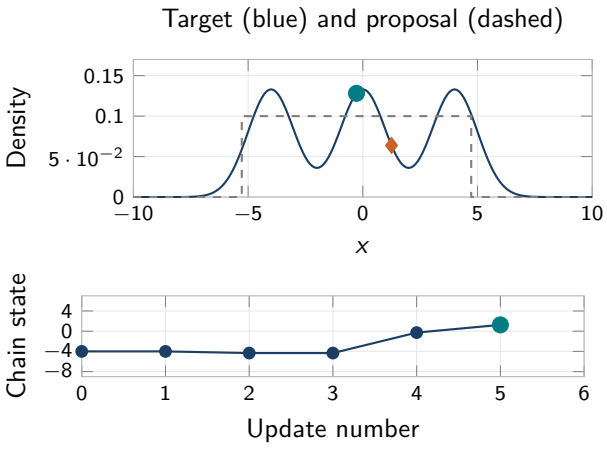
Will the chain move, or repeat its current value?

Advance once to reveal the decision.

An evolving example: propose, decide, record

Update 5 of 6: Decision

$X_4 = -0.2759$



Uniform half-width $h = 5$. Candidate $Y = 1.2520$.

$$\log r = \log \frac{\pi(Y)}{\pi(X_t)} = -0.6978.$$

$$\alpha = \min\{1, e^{-0.6978}\} = 0.4977.$$

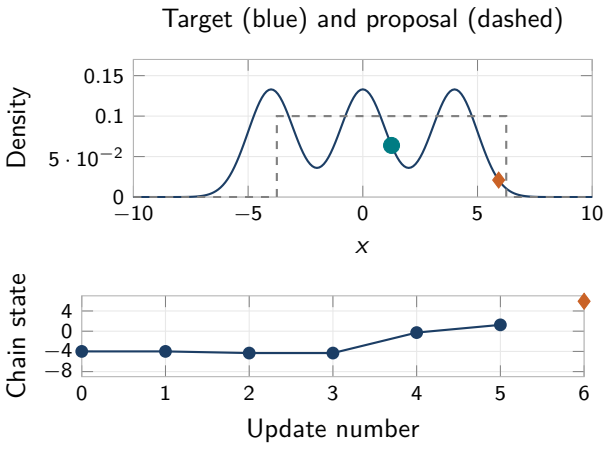
$$U = 0.2983 < \alpha.$$

Accept: $X_5 = 1.2520$.
Append the new value to the chain.

An evolving example: propose, decide, record

Update 6 of 6: Proposal

$X_5 = 1.2520$



Uniform half-width $h = 5$. Candidate $Y = 5.9216$.

The orange diamond is a candidate. It has not yet entered the chain.

$$\alpha(X_t, Y) = \min\{1, \pi(Y)/\pi(X_t)\}.$$

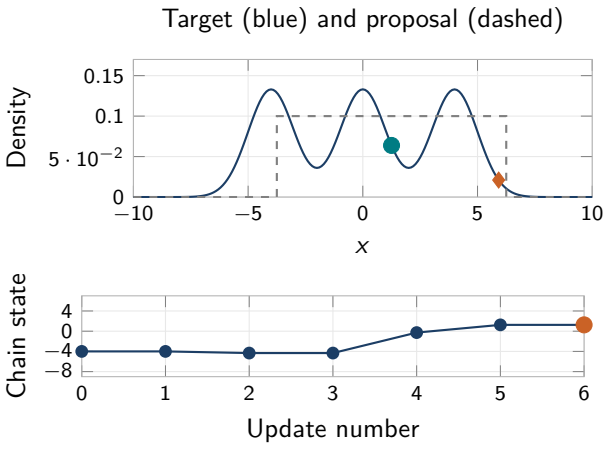
Will the chain move, or repeat its current value?

Advance once to reveal the decision.

An evolving example: propose, decide, record

Update 6 of 6: Decision

$X_5 = 1.2520$



Uniform half-width $h = 5$. Candidate $Y = 5.9216$.

$$\log r = \log \frac{\pi(Y)}{\pi(X_t)} = -1.1115.$$

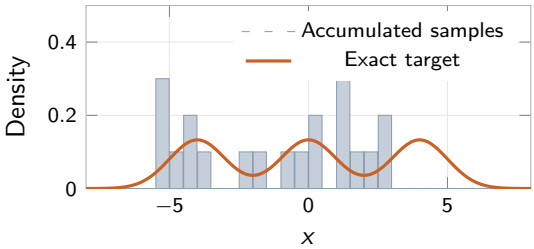
$$\alpha = \min\{1, e^{-1.1115}\} = 0.3291.$$

$$U = 0.9199 \geq \alpha.$$

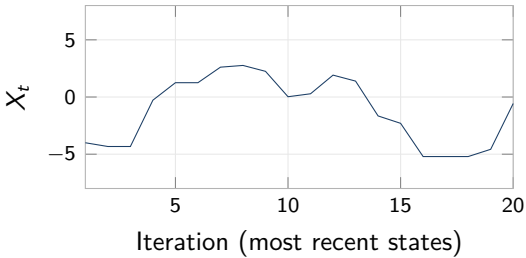
Reject: $X_6 = 1.2520$.
Append the current value again. The repeated point is a sample.

Continuing the same chain: a growing sample distribution

After 20 updates



Including every repeated state.

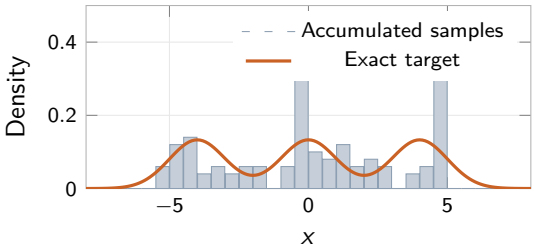


$$\widehat{\mathcal{I}}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = -1.1954 \quad (\mathcal{I} = 0), \qquad \text{acceptance} = 75.0\%.$$

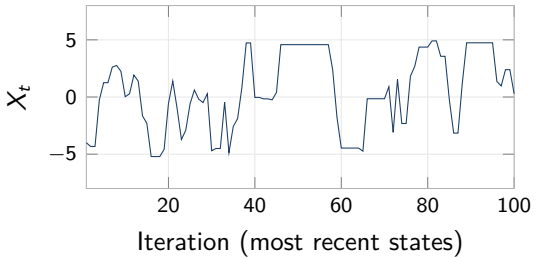
Fixed histogram bins and axes. The histogram uses all n states; the trace shows the latest $\min(n, 100)$. Accuracy need not improve at every checkpoint.

Continuing the same chain: a growing sample distribution

After 100 updates



Including every repeated state.

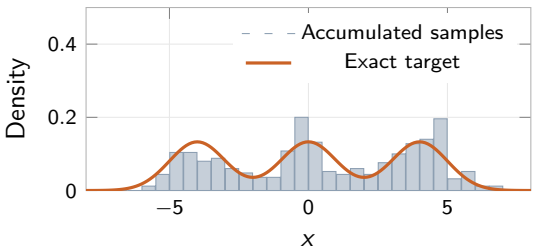


$$\widehat{\mathcal{I}}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = 0.4832 \quad (\mathcal{I} = 0), \qquad \text{acceptance} = 60.0\%.$$

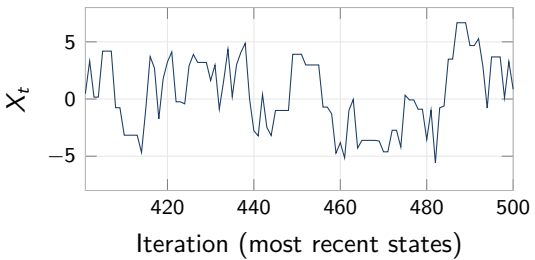
Fixed histogram bins and axes. The histogram uses all n states; the trace shows the latest $\min(n, 100)$. Accuracy need not improve at every checkpoint.

Continuing the same chain: a growing sample distribution

After 500 updates



Including every repeated state.

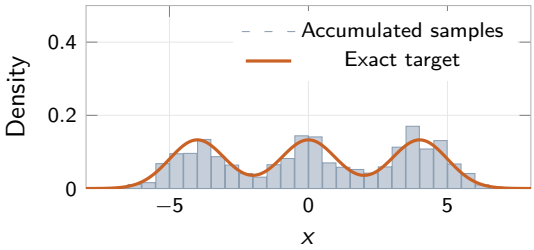


$$\widehat{\mathcal{I}}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = 0.5899 \quad (\mathcal{I} = 0), \qquad \text{acceptance} = 59.0\%.$$

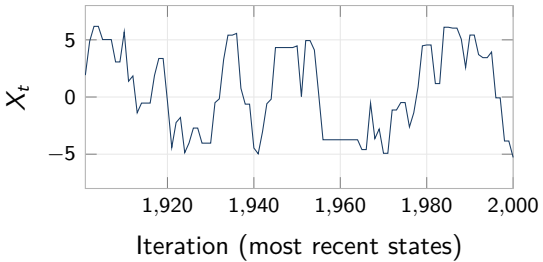
Fixed histogram bins and axes. The histogram uses all n states; the trace shows the latest $\min(n, 100)$. Accuracy need not improve at every checkpoint.

Continuing the same chain: a growing sample distribution

After 2,000 updates



Including every repeated state.

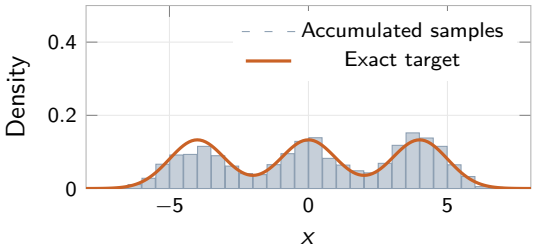


$$\widehat{\mathcal{I}}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = 0.3150 \quad (\mathcal{I} = 0), \qquad \text{acceptance} = 60.8\%.$$

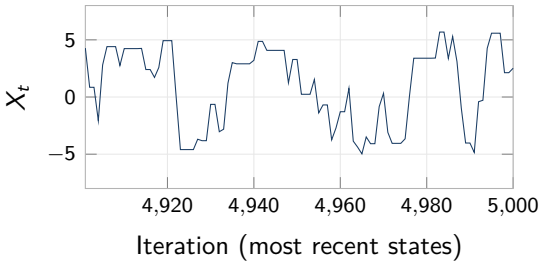
Fixed histogram bins and axes. The histogram uses all n states; the trace shows the latest $\min(n, 100)$. Accuracy need not improve at every checkpoint.

Continuing the same chain: a growing sample distribution

After 5,000 updates



Including every repeated state.



$$\widehat{\mathcal{I}}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = 0.2903 \quad (\mathcal{I} = 0), \qquad \text{acceptance} = 60.4\%.$$

Fixed histogram bins and axes. The histogram uses all n states; the trace shows the latest $\min(n, 100)$. Accuracy need not improve at every checkpoint.

A controlled comparison of proposal scales

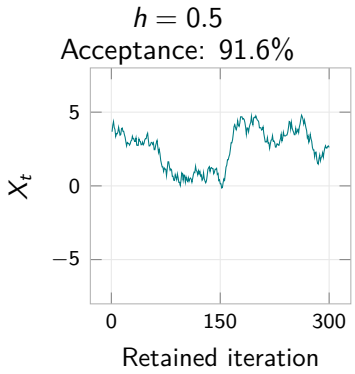
Compare uniform random walks on the same three-component target:

$$Y \mid X_t = x \sim \mathrm{Unif}(x - h, x + h), \qquad h \in \{0.5, 5, 30\}.$$

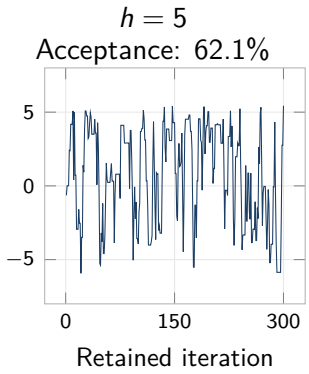
- Each chain starts at $X_0 = -4$.
- Discard the first $B = 2{,}000$ updates.
- Retain $N = 20{,}000$ subsequent states, including repeats.

Compare acceptance, trace plots, autocorrelation and moment estimates. The proposal standard deviation is $h/\sqrt{3}$. These are individual reproducible runs, so numerical values vary with the seed.

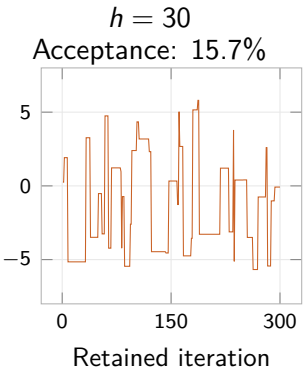
Small steps and large steps both slow exploration



Frequent short moves; crossings between modes are slow.

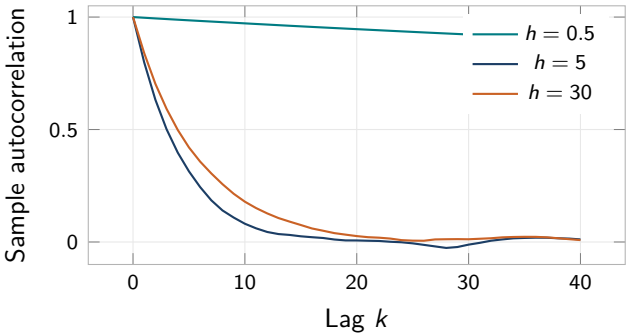


Substantial movement across the target.



Many rejections produce long flat stretches.

Autocorrelation measures dependence across iterations



At stationarity,

$$\rho_k = \mathrm{Corr}(X_t, X_{t+k}).$$

If ρ_k stays large, nearby states supply similar information.
A high acceptance rate alone does not imply an efficient estimate.

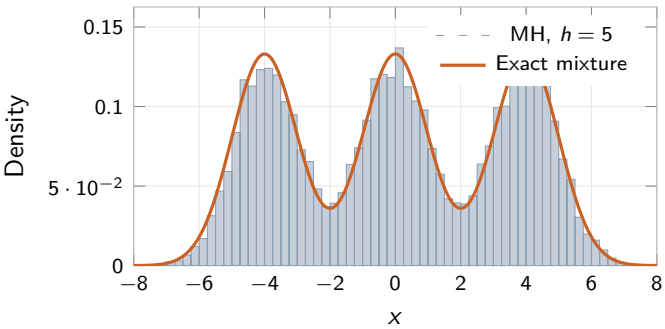
The three-component mixture: numerical results

h	Acceptance	$\widehat{\mathbb{E}[X]}$	$\widehat{\mathbb{E}[X^2]}$	MCSE of mean
0.5	91.6%	0.6074	12.5658	0.2375
5	62.1%	0.1175	11.4773	0.0645
30	15.7%	-0.0683	11.8779	0.0845
Exact	—	0	$35/3$	—

MCSE estimates use 200 consecutive batches of 100 retained states. For very slow mixing, batches of length 100 may still be dependent and underestimate uncertainty.

The middle proposal scale performs best in this comparison. There is no universally best scale or acceptance rate.

A histogram checks the marginal shape



The histogram uses all 20,000 retained states.
Agreement with the target supports the implementation.
A histogram hides the order of observations. Use a trace plot and dependence diagnostics as well.

Monte Carlo error for dependent samples

For a stationary chain with summable autocorrelations of $f(X_t)$,

$$\operatorname{Var}(\widehat{\mathcal{I}}_N) \approx \frac{\operatorname{Var}_\pi(f)}{N} \underbrace{\left(1 + 2\sum_{k=1}^{\infty} \rho_k^{(f)}\right)}_{\tau_f}.$$

The effective sample size is approximately $N_{\text{eff}} = N/\tau_f$.

With batch length $b = 100$ and $m = N/b = 200$, define

$$\bar{f}_j = \frac{1}{b} \sum_{t=(j-1)b+1}^{jb} f(X_t), \qquad \widehat{\mathrm{MCSE}} = \sqrt{\frac{s_{\bar{f}}^2}{m}}.$$

$s_{\bar{f}}^2$ is the sample variance of the m batch means.
Batches should be long enough that their means are approximately independent. MCSE concerns simulation error, not uncertainty in the target distribution.

Example 2: the three-Gaussian target in two dimensions

For $\boldsymbol{x} = (x_1, x_2)^\top \in \mathbb{R}^2$,

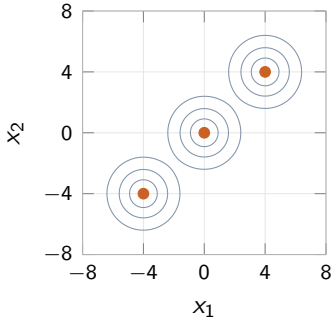
$$\pi_2(\boldsymbol{x}) = \frac{1}{3} \sum_{k=1}^{3} \varphi_2(\boldsymbol{x}; \boldsymbol{m}_k, I_2),$$

$$\boldsymbol{m}_1 = \begin{pmatrix} -4 \\ -4 \end{pmatrix}, \quad \boldsymbol{m}_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{m}_3 = \begin{pmatrix} 4 \\ 4 \end{pmatrix}.$$

Each component has two independent, unit-variance coordinates. Both coordinate marginals equal the target in Example 1.

The mixture creates dependence

The coordinates share one component label. Their joint density is not the product of their marginals.



Exact joint-density contours; dots mark the three component centres.

A uniform proposal over a moving square

Draw $V_1, V_2 \overset{\text{iid}}{\sim} \text{Unif}(-1, 1)$:

$$\boldsymbol{Y} = \boldsymbol{x} + h(V_1, V_2)^\top, \qquad h = 5.$$

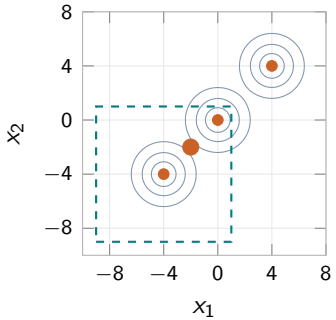
The proposal density is

$$q(\boldsymbol{y} \mid \boldsymbol{x}) = \frac{1}{4h^2} \mathbf{1}\{\|\boldsymbol{y} - \boldsymbol{x}\|_\infty \le h\}.$$

It is symmetric, so

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\{1, \pi_2(\boldsymbol{y})/\pi_2(\boldsymbol{x})\}.$$

Draw one independent $U \sim \text{Unif}(0, 1)$. Accept or reject **both coordinates together**.



At $\boldsymbol{x} = (-4, -4)$ the square is $[-9, 1]^2$; density $1/100$ inside. Orange point at $(-2, -2)$: proposal.

A complete short chain of two-dimensional states

Start $\boldsymbol{X}_0 = (-4, -4)$, with $h = 5$. All proposed moves lie in the current square.

Update	Current	Proposal	α	U	Decision	Next state
1	$(-4,-4)$	$(-2,-2)$	0.0366	0.10	Reject	$(-4,-4)$
2	$(-4,-4)$	$(0,0)$	1.0000	0.40	Accept	$(0,0)$
3	$(0,0)$	$(2,1)$	0.0836	0.02	Accept	$(2,1)$
4	$(2,1)$	$(4,4)$	1.0000	0.60	Accept	$(4,4)$
5	$(4,4)$	$(6,4)$	0.1353	0.20	Reject	$(4,4)$
6	$(4,4)$	$(3,3)$	0.3680	0.30	Accept	$(3,3)$

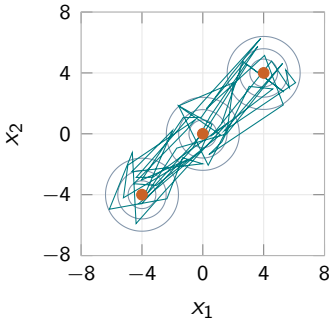
For the first proposal,

$$r = \frac{2e^{-4} + e^{-36}}{1 + e^{-16} + e^{-64}} \approx 0.0366.$$

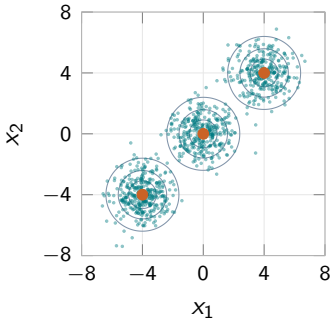
Since $0.10 > 0.0366$, reject and record $(-4,-4)$ again. The third update accepts $(2,1)$: the chain need not stay on the diagonal.

Exploring the joint distribution

First 300 retained states



Scatter across the retained run



Start $(-4, -4)$, $h = 5$. Discard 2,000 updates; retain 20,000 states. The scatter displays every 20th state for readability; all states enter estimates. Contours show the exact target. Rejections repeat the whole current vector.