Why construct a Markov chain?

We want to sample from a target density $\pi(x)$ on $\mathbb{R}^d$. Direct sampling may be difficult, even when we can evaluate

$$\pi(x) = c\, g(x), \qquad g(x) \text{ known}, \quad c \text{ unknown}.$$

The MCMC idea

Construct a Markov chain whose invariant distribution is π. After suitable convergence, use its visited states for Monte Carlo estimation.

For example,

$$\mathcal{I} = \mathbb{E}_\pi[f(X)], \qquad \widehat{\mathcal{I}}_N = \frac{1}{N} \sum_{t=B+1}^{B+N} f(x_t).$$

Here B states are discarded initially; the retained states are generally dependent.

Recall: invariance and convergence

For a finite state space, T_{ij} is the probability of moving from state i to state j.

$$\pi T = \pi$$

means that a chain started with distribution π keeps that distribution.

- **Irreducibility:** the chain can reach all relevant parts of its state space.
- **Aperiodicity:** the chain is not forced into a fixed cycle.
- **Invariance:** the update preserves the desired target.

On a finite state space, irreducibility and aperiodicity give convergence to the unique invariant distribution. For continuous spaces, corresponding recurrence and ergodicity conditions are needed.

Detailed balance implies invariance

The detailed-balance condition is

$$\pi_i T_{ij} = \pi_j T_{ji}.$$

At equilibrium, probability flow from i to j equals the reverse flow. Summing over i,

$$\sum_i \pi_i T_{ij} = \pi_j \sum_i T_{ji} = \pi_j.$$

Thus $\pi T = \pi$.

What this establishes

Detailed balance is sufficient for invariance. It does not, by itself, establish convergence from every starting state.

For continuous spaces, the corresponding identity is expressed using transition kernels and integrals.

The Metropolis–Hastings update

Choose x_0 with $\pi(x_0) > 0$. For $t = 0, \ldots, T-1$:

1. Propose $x^* \sim q(\cdot \mid x_t)$.
2. Compute
$$a(x^*, x_t) = \min\left\{1, \frac{\pi(x^*)}{\pi(x_t)} \frac{q(x_t \mid x^*)}{q(x^* \mid x_t)}\right\}.$$
3. Draw $u \sim U(0,1)$, independently of the proposal.
4. If $u < a$, set $x_{t+1} = x^*$; otherwise set $x_{t+1} = x_t$.

The proposal supplies a candidate; the acceptance rule corrects its distribution.

A rejection is still an iteration

Suppose three proposals produce the decisions accept, reject, accept:

$$x_0 \longrightarrow x_1 \longrightarrow x_1 \longrightarrow x_3.$$

The repeated value must appear in the stored chain.

- An accepted proposal changes the state to the candidate.
- A rejected proposal leaves the state where it is.
- Store a state after *every* iteration, including rejections.

Do not retain only accepted moves

Removing repeated states generally changes the empirical distribution. The time spent at each state is part of the MH construction.

The transition includes a probability of staying put

For a continuous proposal, let

$$r(x) = 1 - \int q(y \mid x) a(y, x)\, dy.$$

The complete transition kernel is

$$K(x, dy) = q(y \mid x) a(y, x)\, dy + r(x) \delta_x(dy).$$

- The first term describes accepted proposals.
- The second term is a point mass at the current state, caused by rejection.

For distinct states, the moving part has density $q(y \mid x) a(y, x)$. The diagonal part must also be retained in the complete transition.

Why MH works: the forward probability flow

Write the current state as x and the candidate as y. Then

$$a(y, x) = \min\left\{1, \frac{\pi(y)q(x \mid y)}{\pi(x)q(y \mid x)}\right\}.$$

Multiply by the equilibrium density and the proposal density:

$$\begin{aligned}\pi(x)q(y \mid x)a(y, x) &= \pi(x)q(y \mid x)\min\left\{1, \frac{\pi(y)q(x \mid y)}{\pi(x)q(y \mid x)}\right\} \\ &= \min\{\pi(x)q(y \mid x),\ \pi(y)q(x \mid y)\}.\end{aligned}$$

The expression is symmetric in x and y.

Why MH works: the reverse probability flow

Interchanging x and y gives

$$\pi(y)q(x \mid y)a(x,y) = \min\{\pi(y)q(x \mid y),\ \pi(x)q(y \mid x)\}.$$

Therefore

$$\boxed{\pi(x)q(y \mid x)a(y,x) = \pi(y)q(x \mid y)a(x,y).}$$

The stay-put term lies on the diagonal and also satisfies detailed balance.

Consequence

The full MH kernel preserves π. With suitable ergodicity conditions, the chain can be used to estimate expectations under π.

Why an unknown normalising constant is harmless

If $\pi(x) = c\, g(x)$, then

$$\frac{\pi(x^*)}{\pi(x_t)} = \frac{c\, g(x^*)}{c\, g(x_t)} = \frac{g(x^*)}{g(x_t)}.$$

Hence

$$a = \min\left\{1, \frac{g(x^*)q(x_t \mid x^*)}{g(x_t)q(x^* \mid x_t)}\right\}.$$

- We need target density ratios, not the value of c.
- A factor may be dropped only if it is constant with respect to the sampled state.
- If a candidate is outside the target support, its target density is zero and it is rejected.

Two random-walk proposals

Gaussian random walk:

$$x^* = x_t + \sigma Z, \qquad Z \sim N(0, I_d),$$

so $q(\cdot \mid x_t) = N(x_t, \sigma^2 I_d)$.

Uniform random walk:

$$x^* = x_t + \sigma V, \qquad V_j \overset{\text{iid}}{\sim} U(-1, 1).$$

The candidate is uniform on the hypercube

$$\prod_{j=1}^{d} [x_{t,j} - \sigma, x_{t,j} + \sigma].$$

In both cases $\sigma > 0$ controls the proposal scale, but the uniform increment has variance $\sigma^2/3$ per coordinate.

Symmetry simplifies the acceptance probability

Both random-walk proposals satisfy

$$q(x^* \mid x_t) = q(x_t \mid x^*).$$

For the uniform proposal, the density is $(2\sigma)^{-d}$ whenever $|x_j^* - x_{t,j}| \le \sigma$ for all j. This condition is unchanged by reversing the move. Thus

$$\boxed{a = \min\left\{1, \frac{g(x^*)}{g(x_t)}\right\}.}$$

- Higher-density candidates are accepted with probability one.
- Lower-density candidates can still be accepted.

Allowing moves to lower-density regions is essential for sampling a distribution.

Implement the acceptance test on the log scale

Set $\ell(x) = \log g(x)$ and compute

$$L = \ell(x^*) - \ell(x_t) + \log q(x_t \mid x^*) - \log q(x^* \mid x_t).$$

Then accept if

$$\boxed{\log u < \min\{0, L\}, \qquad u \sim U(0,1).}$$

For a symmetric proposal,

$$L = \ell(x^*) - \ell(x_t).$$

- This avoids exponentiating very large or small numbers.
- Assign $\ell(x^*) = -\infty$ outside the target support.
- Keep the current state when the test fails.

Example 1: standard Gaussian target

Target:

$$\pi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

Suppose we use only $g(x) = e^{-x^2/2}$. Take the uniform random walk with $\sigma = 0.5$:

$$x^* \mid x_t \sim U(x_t - 0.5, x_t + 0.5).$$

The acceptance probability is

$$\boxed{a = \min\{1, \exp[-\tfrac{1}{2}((x^*)^2 - x_t^2)]\}.}$$

Equivalently, $L = -\frac{1}{2}((x^*)^2 - x_t^2)$.

Example 1: three explicit acceptance decisions

Start at $x_0 = 1$. Each candidate lies within 0.5 of the current state.

t	x_t	x^*	L	a	u	x_{t+1}
0	1.0	0.7	0.255	1.000	0.60	0.7 (accept)
1	0.7	1.1	−0.360	0.698	0.80	0.7 (reject)
2	0.7	1.0	−0.255	0.775	0.50	1.0 (accept)

For the rejected move,

$$L = -\tfrac{1}{2}(1.1^2 - 0.7^2) = -0.36, \qquad a = e^{-0.36} \approx 0.698.$$

These chosen random draws illustrate the rule; the stored sequence is

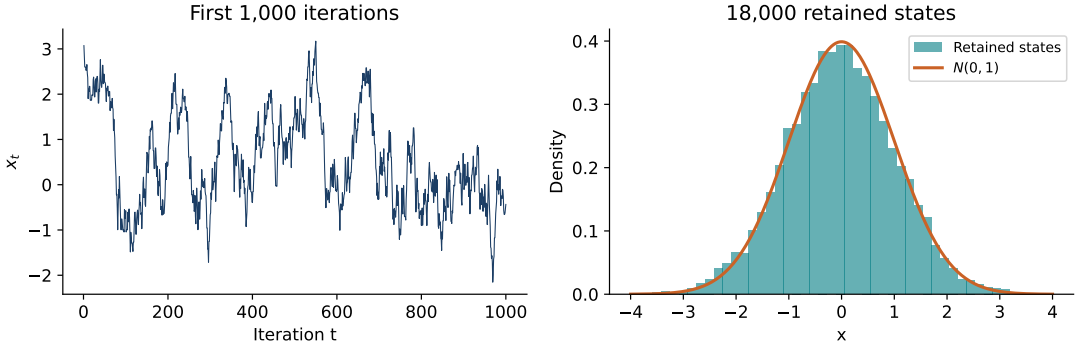
$$1.0,\ 0.7,\ 0.7,\ 1.0.$$

Example 1: MATLAB implementation

```
rng(7201); T = 20000; sigma = 0.5;
x = zeros(T+1,1); x(1) = 3;
for t = 1:T
    candidate = x(t) + sigma*(2*rand-1);
    logRatio = -0.5*(candidate^2-x(t)^2);
    x(t+1) = x(t);          % retain on rejection
    if log(rand) < min(0,logRatio)
        x(t+1) = candidate;
    end
end
B = 2000;
kept = x(B+2:end);          % states x_(B+1),...,x_T
```

MATLAB index $t+1$ stores mathematical state x_t; index 1 stores x_0.

Example 1: trace and empirical density



Uniform half-width 0.5; $x_0 = 3$; 20,000 iterations; first 2,000 discarded. The histogram retains repeated states. Agreement in this run is a useful check, not a general convergence guarantee.

Example 2: a curved two-dimensional target

Write the state as $z = (x, y)$ and take

$$\pi(x, y) = c \exp(-(1 - x)^2 - 10(y - x^2)^2).$$

We evaluate only

$$\ell(x, y) = -(1 - x)^2 - 10(y - x^2)^2.$$

- $(1 - x)^2$ concentrates the first coordinate around 1.
- $10(y - x^2)^2$ concentrates the second coordinate near the curve $y = x^2$.
- The high-density region is curved and narrow.

A Gaussian random walk must explore both coordinates without knowing c.

Example 2: Gaussian proposal and acceptance rule

Propose both coordinates together:

$$z^* = z_t + \sigma\xi, \qquad \xi \sim N(0, I_2).$$

Thus $q(\cdot \mid z_t) = N(z_t, \sigma^2 I_2)$ and the proposal is symmetric.

$$\begin{aligned} L &= \ell(x^*, y^*) - \ell(x_t, y_t) \\ &= -(1 - x^*)^2 - 10(y^* - (x^*)^2)^2 \\ &\quad + (1 - x_t)^2 + 10(y_t - x_t^2)^2. \end{aligned}$$

Accept the *whole candidate vector* if $\log u < \min(0, L)$. Otherwise keep both current coordinates.

Example 2: one accepted and one rejected move

Start at $z_0 = (1, 1)$, where $\ell(z_0) = 0$. Take $\sigma = 0.5$.

1. A proposal $z^* = (1.2, 1.4)$ has

$$\ell(z^*) = -0.2^2 - 10(1.4 - 1.44)^2 = -0.056.$$

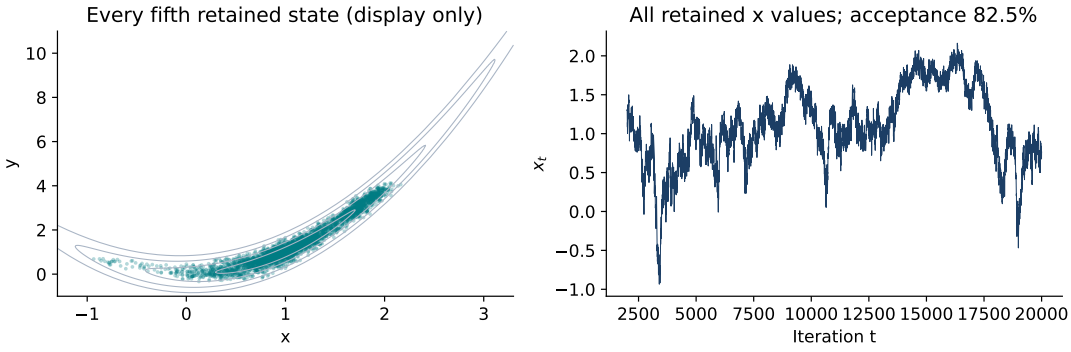
   Thus $a = e^{-0.056} \approx 0.946$. With $u = 0.40$, accept: $z_1 = (1.2, 1.4)$.

2. Next propose $(1.4, 1.4)$, whose log density is

$$-0.4^2 - 10(1.4 - 1.96)^2 = -3.296.$$

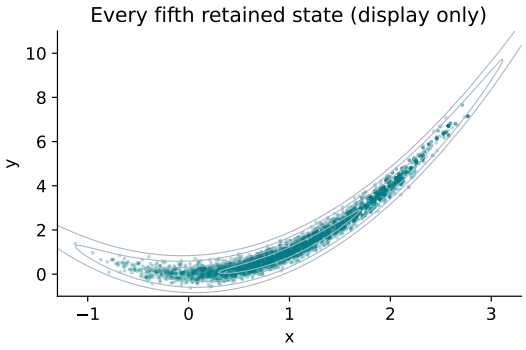
   Now $L = -3.296 - (-0.056) = -3.240$, so $a \approx 0.039$. With $u = 0.20$, reject and retain $z_2 = (1.2, 1.4)$.

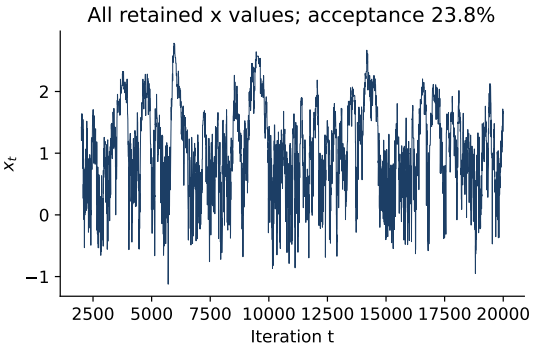
Example 2: small proposal scale, $\sigma = 0.05$



20,000 iterations from $(1, 1)$; first 2,000 discarded. Small moves can be accepted frequently while the chain explores slowly. Acceptance alone misses this problem.

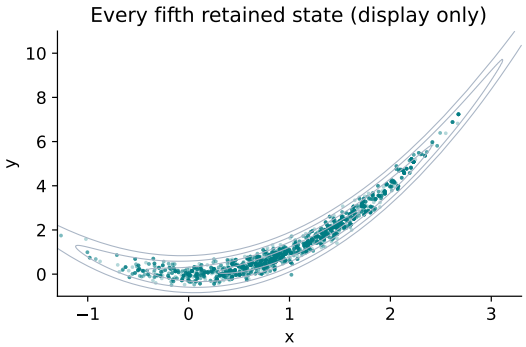
Example 2: intermediate proposal scale, $\sigma = 0.5$

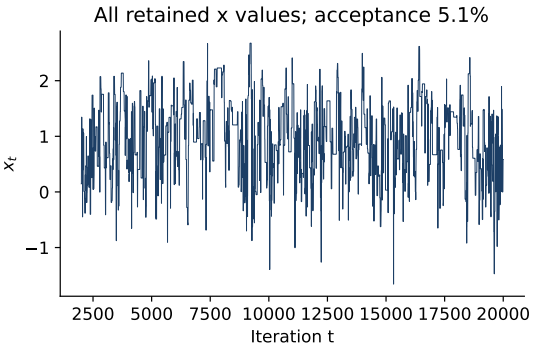




Same run length and initial state. Compare the range visited and the persistence of the trace with the small-scale run. This is a finite-run comparison, not a claim of optimal tuning.

Example 2: large proposal scale, $\sigma = 2$





Large isotropic proposals often land outside the narrow high-density region. Rejections create repeated states; occasional larger moves do not guarantee efficient exploration.

Choosing the random-walk proposal scale

Scale	Typical behaviour	Main concern
Too small	Many accepted moves	Moves are short; strong dependence
Too large	Many rejected moves	Long runs of repeated states
Intermediate	Useful accepted moves	Assess exploration as well as acceptance

There is no universal best acceptance rate

Performance depends on the target geometry, proposal, dimension and quantity being estimated. For random-walk MH, a very high acceptance rate can indicate an overly small scale.

Using the output responsibly

- Inspect traces for drift, persistent dependence and long repeated stretches.
- Compare chains from dispersed starting points when feasible.
- Treat discarding an initial segment as a practical choice, not proof of convergence.
- Estimate $\mathcal{I}$ using all retained states, including repetitions.

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

Sample dependence affects Monte Carlo uncertainty. A smooth histogram or a high acceptance rate alone cannot establish reliable exploration.