Metropolis–Hastings in Practice

Warm-up: what do we need for convergence to the posterior?

We want an MCMC chain with states $\theta^{(0)}, \theta^{(1)}, \ldots$ whose long-run histogram matches the posterior $\pi(\theta \mid y)$.

Three key conditions (finite-state intuition, but same spirit more generally):

1. **Aperiodic:** the chain does not get trapped in a rigid cycle (e.g. bouncing deterministically between two states).
2. **Irreducible:** from any current state, there is a (possibly multi-step) positive probability of reaching any region/state we care about.
3. **Stationarity / invariance:** the transition rule P preserves the target:

$$\pi P = \pi,$$

i.e. if $\theta^{(0)} \sim \pi$, then $\theta^{(1)} \sim \pi$ as well.

Interpretation: (1) avoids systematic cycling, (2) ensures exploration, (3) ensures we converge to the *right* distribution.

warm-up: why we build MCMC chains

Often the posterior $\pi(\theta \mid y)$ is hard to be sampled from directly.

$$\pi(\theta \mid y) \;\propto\; \underbrace{\pi(y \mid \theta)}_{\text{likelihood}}\,\underbrace{\pi(\theta)}_{\text{prior}} \quad \text{but the normalising constant is unknown.}$$

Goal: generate samples $\theta^{(1)}, \ldots, \theta^{(N)}$ so we can compute:

posterior mean, median, mode, variance, credible intervals, probabilities, ...

Mechanism: start somewhere, propose a move, accept/reject, repeat:

$$\theta^{(i-1)} \to \theta' \to \begin{cases} \theta^{(i)} = \theta' & \text{(accept)} \\ \theta^{(i)} = \theta^{(i-1)} & \text{(reject)} \end{cases}$$

If the conditions hold, then the long-run empirical distribution of visited states approximates $\pi(\theta \mid y)$.

Metropolis–Hastings recap: the acceptance probability

We target $\pi(\theta \mid y)$ using a proposal distribution $q(\theta' \mid \theta)$.

Generic MH acceptance probability:

$$p_{\text{acc}}(\theta, \theta') = \min\left\{1,\ \frac{\pi(\theta' \mid y)}{\pi(\theta \mid y)} \cdot \frac{q(\theta \mid \theta')}{q(\theta' \mid \theta)}\right\}.$$

Symmetric random walk proposals: if $q(\theta' \mid \theta) = q(\theta \mid \theta')$, then the proposal ratio cancels:

$$p_{\text{acc}}(\theta, \theta') = \min\left\{1,\ \frac{\pi(\theta' \mid y)}{\pi(\theta \mid y)}\right\}.$$

Implementation tip: compute on the log-scale to avoid underflow:

$$\log p^{\star}_{\text{acc}} = \log \pi(\theta' \mid y) - \log \pi(\theta \mid y) + \log q(\theta \mid \theta') - \log q(\theta' \mid \theta),$$

then accept if $\log U \le \log p^{\star}_{\text{acc}}$, $U \sim \text{Unif}(0, 1)$.

Detailed balance: the easy sufficient condition for stationarity

A common way to guarantee $\pi P = \pi$ is **detailed balance**.

Detailed balance equation:

$$\pi(i)\, p_{ij} = \pi(j)\, p_{ji} \quad \text{for all states } i, j.$$

Intuition: in equilibrium, probability *flow* from $i \to j$ matches the flow from $j \to i$.

Key fact: Metropolis–Hastings constructs a transition rule that satisfies detailed balance with respect to $\pi(\cdot \mid y)$. Therefore π is stationary for the MH Markov chain.

So: if the chain is also irreducible and aperiodic, it converges to $\pi(\theta \mid y)$.

Back to theory: why does Metropolis–Hastings work?

Some of you may be thinking: how do we know Metropolis–Hastings gives us the stationary distribution we want?

Fortunately, we spent the last week building theory:

- when a Markov chain converges to a stationary distribution,
- and how to prove stationarity.

Back to theory: why does Metropolis–Hastings work?

Some of you may be thinking: how do we know Metropolis–Hastings gives us the stationary distribution we want?

Fortunately, we spent the last week building theory:

- when a Markov chain converges to a stationary distribution,
- and how to prove stationarity.

Now we apply that theory to MH.

Theorem 5.6: the Markov chain generated by Metropolis–Hastings satisfies detailed balance with respect to the posterior distribution.

So the stationary distribution is exactly the posterior we care about.

Theorem 5.6: set up the detailed balance statement

Let the current state be θ, and the proposed state be θ'.

Detailed balance (what we want to show) is:

$$\pi(\theta \mid y)\, P(\theta \to \theta') = \pi(\theta' \mid y)\, P(\theta' \to \theta).$$

In MH, the move probability factors into:

$$P(\theta \to \theta') = q(\theta' \mid \theta)\, a(\theta, \theta'),$$

where q is the proposal density and a is the acceptance probability.

Theorem 5.6: set up the detailed balance statement

Let the current state be θ, and the proposed state be θ'.

Detailed balance (what we want to show) is:

$$\pi(\theta \mid y)\, P(\theta \to \theta') = \pi(\theta' \mid y)\, P(\theta' \to \theta).$$

In MH, the move probability factors into:

$$P(\theta \to \theta') = q(\theta' \mid \theta)\, a(\theta, \theta'),$$

where q is the proposal density and a is the acceptance probability.

So we want to show:

$$\pi(\theta \mid y)\, q(\theta' \mid \theta)\, a(\theta, \theta') = \pi(\theta' \mid y)\, q(\theta \mid \theta')\, a(\theta', \theta).$$

Theorem 5.6: plug in the MH acceptance rule

MH acceptance probability:

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

Theorem 5.6: plug in the MH acceptance rule

MH acceptance probability:

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

Multiply by $\pi(\theta \mid y)\, q(\theta' \mid \theta)$:

$$\pi(\theta \mid y)\, q(\theta' \mid \theta)\, a(\theta, \theta') = \min\Big(\pi(\theta \mid y)\, q(\theta' \mid \theta),\ \pi(\theta' \mid y)\, q(\theta \mid \theta')\Big).$$

Theorem 5.6: plug in the MH acceptance rule

MH acceptance probability:

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

Multiply by $\pi(\theta \mid y)\, q(\theta' \mid \theta)$:

$$\pi(\theta \mid y)\, q(\theta' \mid \theta)\, a(\theta, \theta') = \min\Big(\pi(\theta \mid y)\, q(\theta' \mid \theta),\ \pi(\theta' \mid y)\, q(\theta \mid \theta')\Big).$$

Now do the same on the other side:

$$\pi(\theta' \mid y)\, q(\theta \mid \theta')\, a(\theta', \theta) = \min\Big(\pi(\theta' \mid y)\, q(\theta \mid \theta'),\ \pi(\theta \mid y)\, q(\theta' \mid \theta)\Big).$$

Theorem 5.6: plug in the MH acceptance rule

MH acceptance probability:

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

Multiply by $\pi(\theta \mid y)\, q(\theta' \mid \theta)$:

$$\pi(\theta \mid y)\, q(\theta' \mid \theta)\, a(\theta, \theta') = \min\Big(\pi(\theta \mid y)\, q(\theta' \mid \theta),\ \pi(\theta' \mid y)\, q(\theta \mid \theta')\Big).$$

Now do the same on the other side:

$$\pi(\theta' \mid y)\, q(\theta \mid \theta')\, a(\theta', \theta) = \min\Big(\pi(\theta' \mid y)\, q(\theta \mid \theta'),\ \pi(\theta \mid y)\, q(\theta' \mid \theta)\Big).$$

These are the same expression (just written in the opposite order), hence:

$$\pi(\theta \mid y)\, q(\theta' \mid \theta)\, a(\theta, \theta') = \pi(\theta' \mid y)\, q(\theta \mid \theta')\, a(\theta', \theta).$$

So MH satisfies detailed balance w.r.t. the posterior.

What this proof buys us (and what's next)

From Theorem 5.6:

- The MH chain satisfies detailed balance w.r.t. the posterior.
- Therefore the posterior is stationary for the MH chain.
- With the other conditions (irreducibility + aperiodicity), we are guaranteed to converge to the posterior distribution.

What this proof buys us (and what's next)

From Theorem 5.6:

- The MH chain satisfies detailed balance w.r.t. the posterior.
- Therefore the posterior is stationary for the MH chain.
- With the other conditions (irreducibility + aperiodicity), we are guaranteed to converge to the posterior distribution.

So everything is OK: we can use Metropolis–Hastings to sample from posterior distributions.

Example 5.3: reaction-time data (lorry drivers)

Story: reaction times (milliseconds) to an obstacle in a lab setting.

Data: $y_1, \ldots, y_n$ (in the notes; here $n = 30$).

Model:

$$Y_i \mid \mu \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2), \qquad \sigma^2 \text{ known.}$$

Parameter of interest: the population mean reaction time μ.

Important constraint: $\mu > 0$ (negative reaction times make no sense).

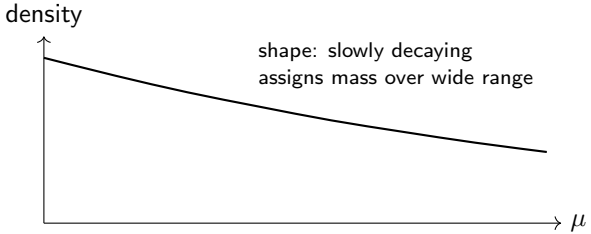
Why not just use conjugacy? Normal prior on μ is conjugate, but assigns positive probability to $\mu < 0$. So we choose a prior supported on $(0, \infty)$.

Example 5.3: prior choice (non-conjugate on purpose)

We choose an exponential prior:

$$\mu \sim \mathrm{Exp}(\lambda), \quad \lambda = 0.01, \qquad p(\mu) = \lambda e^{-\lambda \mu}\, \mathbf{1}_{\{\mu \geq 0\}}.$$

Interpretation: $\lambda = 0.01$ is a *very flat* exponential on the scale of tens/hundreds (in ms), so it is fairly weakly informative.



Consequence: posterior is *not* a standard named distribution $\Rightarrow$ use MCMC.

Example 5.3: posterior (up to proportionality)

Likelihood:

$$p(y \mid \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i-\mu)^2}{2\sigma^2}\right) \;\propto\; \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right).$$

Prior:

$$p(\mu) \propto \exp(-\lambda\mu)\mathbf{1}_{\{\mu\geq 0\}}, \quad \lambda = 0.01.$$

Posterior (kernel):

$$\pi(\mu \mid y) \;\propto\; \exp\left(-\lambda\mu - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right)\mathbf{1}_{\{\mu\geq 0\}}.$$

Key point: multiplying exponential prior and Gaussian likelihood does *not* yield a convenient closed-form posterior family here (with the positivity constraint). So we sample using random-walk Metropolis–Hastings.

Example 5.3: proposal (uniform random walk) + symmetry

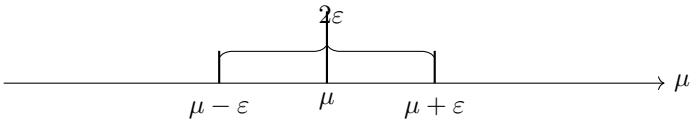
We use a random-walk proposal:

$$\mu' \mid \mu \sim \mathrm{Unif}(\mu - \varepsilon,\ \mu + \varepsilon).$$

Symmetry: for any μ, μ' in range,

$$q(\mu' \mid \mu) = \frac{1}{2\varepsilon} \quad \text{and} \quad q(\mu \mid \mu') = \frac{1}{2\varepsilon} \quad \Rightarrow \quad \frac{q(\mu \mid \mu')}{q(\mu' \mid \mu)} = 1.$$

Also respect the support: if $\mu' < 0$, reject immediately (posterior is zero there).



Example 5.3: acceptance probability (posterior ratio)

Because the proposal is symmetric, MH acceptance is:

$$p_{\text{acc}}(\mu, \mu') = \min\left\{1,\ \frac{\pi(\mu' \mid y)}{\pi(\mu \mid y)}\right\}.$$

Using the posterior kernel:

$$\frac{\pi(\mu' \mid y)}{\pi(\mu \mid y)} = \exp\left(-\lambda(\mu' - \mu) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(y_i - \mu')^2 - (y_i - \mu)^2\right]\right),$$

provided $\mu' \geq 0$; otherwise the ratio is 0.

Log acceptance (recommended in code):

$$\log p^{\star}_{\text{acc}} = -\lambda\mu' - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu')^2 \ - \ \Big(-\lambda\mu - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\Big).$$

Then accept if $\log U \leq \log p^{\star}_{\text{acc}}$, $U \sim \text{Unif}(0,1)$.

Example 5.3: MH algorithm (step-by-step)

Inputs: data $y_{1:n}$, known σ^2, step size ε, iterations T.

Algorithm (random-walk MH on $\mu > 0$):

1. Initialise: choose $\mu^{(0)} > 0$ (e.g. $\mu^{(0)} = 1$).
2. For $t = 1, 2, \ldots, T$:
   1. Propose $\mu' \sim \mathrm{Unif}(\mu^{(t-1)} - \varepsilon,\ \mu^{(t-1)} + \varepsilon)$.
   2. If $\mu' < 0$: reject, set $\mu^{(t)} = \mu^{(t-1)}$, continue.
   3. Compute $\log p^{\star}_{\mathrm{acc}} = \log \pi(\mu' \mid y) - \log \pi(\mu^{(t-1)} \mid y)$.
   4. Sample $U \sim \mathrm{Unif}(0, 1)$. If $\log U \leq \log p^{\star}_{\mathrm{acc}}$, accept: $\mu^{(t)} = \mu'$; else reject: $\mu^{(t)} = \mu^{(t-1)}$.

Output: the chain $\{\mu^{(t)}\}_{t=0}^{T}$, whose post-burn-in histogram approximates $\pi(\mu \mid y)$.

Example 5.3: diagnostics (trace plot, burn-in, and the caterpillar)

Trace plot: plot $\mu^{(t)}$ versus iteration t.

- Early iterations may show a drift from the initial value into the high-posterior region.
- After convergence, we want a "fat hairy caterpillar": lots of local movement, no obvious trend.

Burn-in: discard first B iterations (e.g. $B \approx 300$ in the example).

$$\{\mu^{(B+1)}, \ldots, \mu^{(T)}\} \;\approx\; \text{samples from } \pi(\mu \mid y).$$

Posterior summaries from samples:

$$\widehat{\mathbb{E}[\mu \mid y]} = \frac{1}{T-B} \sum_{t=B+1}^{T} \mu^{(t)}, \qquad \text{CI}_{0.95} = \big[\text{quantile}_{0.025},\ \text{quantile}_{0.975}\big].$$

Important practical note: MH guarantees eventual convergence (under conditions), but does *not* tell you *how long* it takes.

Example 5.3: effect of the prior (informative data case)

In the worked example, the posterior credible interval under the exponential prior is extremely close to what you get under a conjugate normal prior.

Interpretation:

- The data are strongly informative about μ.
- The prior contributes relatively little (posterior dominated by likelihood).

Lesson: using a more realistic prior support ($\mu > 0$) can be conceptually cleaner, and when data are informative it often changes little numerically.

But: in smaller samples or noisier settings, the prior choice can matter a lot more.

Example 5.4: Pareto model (link to Lab 3.9 MAP estimate)

Now we revisit a model where we previously found the MAP (posterior mode), but did not yet have MCMC.

Model: $Y_1, \dots, Y_n \overset{\text{iid}}{\sim} \text{Pareto}(\beta)$ with support $y \geq 1$, density

$$p(y \mid \beta) = \frac{\beta}{y^{\beta+1}}, \qquad y \geq 1,\ \beta > 0.$$

Prior: $\beta \sim \text{Gamma}(a, b)$ with density kernel

$$p(\beta) \propto \beta^{a-1} e^{-b\beta}, \qquad \beta > 0.$$

Goal: sample from $\pi(\beta \mid y)$ using MH.

Example 5.4: posterior kernel

Likelihood:

$$p(y \mid \beta) = \prod_{i=1}^{n} \frac{\beta}{y_i^{\beta+1}} = \beta^n \prod_{i=1}^{n} y_i^{-(\beta+1)}.$$

Prior: $p(\beta) \propto \beta^{a-1} e^{-b\beta}$.

Posterior (up to proportionality):

$$\pi(\beta \mid y) \propto \beta^{n+a-1}\, e^{-b\beta} \prod_{i=1}^{n} y_i^{-(\beta+1)}\, \mathbf{1}_{\{\beta>0\}}.$$

Not a standard family: the $\prod y_i^{-(\beta+1)}$ term prevents a simple gamma posterior. So again: use Metropolis–Hastings.

Example 5.4: proposal and acceptance probability

Proposal (normal random walk):

$$\beta' \mid \beta \sim \mathcal{N}(\beta, \sigma_q^2).$$

This is symmetric: $q(\beta' \mid \beta) = q(\beta \mid \beta')$.

Support constraint: if $\beta' \leq 0$, reject immediately.

Acceptance probability (symmetric proposal):

$$p_{\text{acc}}(\beta, \beta') = \min\left\{1,\ \frac{\pi(\beta' \mid y)}{\pi(\beta \mid y)}\right\}.$$

Plug in the kernel:

$$\frac{\pi(\beta' \mid y)}{\pi(\beta \mid y)} = \left(\frac{\beta'}{\beta}\right)^{n+a-1} \exp\big(-b(\beta' - \beta)\big) \prod_{i=1}^{n} y_i^{-(\beta' - \beta)}.$$

Log acceptance:

$$\log p_{\text{acc}}^{\star} = (n + a - 1)(\log \beta' - \log \beta) - b(\beta' - \beta) - (\beta' - \beta) \sum_{i=1}^{n} \log y_i.$$

Example 5.4: MH algorithm for β

Algorithm:

1. Initialise: choose $\beta^{(0)} > 0$.
2. For $t = 1, \ldots, T$:
   1. Propose $\beta' \sim \mathcal{N}(\beta^{(t-1)}, \sigma_q^2)$.
   2. If $\beta' \leq 0$: reject, set $\beta^{(t)} = \beta^{(t-1)}$.
   3. Else compute $\log p_{\text{acc}}^{\star} = \log \pi(\beta' \mid y) - \log \pi(\beta^{(t-1)} \mid y)$.
   4. Accept if $\log U \leq \log p_{\text{acc}}^{\star}$, $U \sim \text{Unif}(0, 1)$.

Practical note: tune σ_q so the chain moves (not too sticky) but does not reject almost everything. (We will discuss tuning + diagnostics later.)

Mini-exercise slide (in-class)

Quick check (2–4 minutes):

1. Starting from
$$\pi(\beta \mid y) \propto \beta^{n+a-1} e^{-b\beta} \prod_{i=1}^{n} y_i^{-(\beta+1)},$$
derive $\log \pi(\beta \mid y)$ up to an additive constant.
2. Show that the proposal ratio cancels for the normal random walk proposal.
3. Confirm the compact log-acceptance expression:
$$\log p_{\text{acc}}^{\star} = (n + a - 1)(\log \beta' - \log \beta) - b(\beta' - \beta) - (\beta' - \beta) \sum_{i=1}^{n} \log y_i.$$

Why we like this form: in code it is a few stable operations: logs + sums, no giant products.

Closing summary

1) Targeting a posterior without closed form: build a Markov chain whose stationary distribution is $\pi(\cdot \mid y)$.

2) Convergence conditions:

- Aperiodic + irreducible + $\pi P = \pi$ (often via detailed balance) $\Rightarrow$ convergence to π.

3) Metropolis–Hastings template:

- Choose proposal $q(\cdot \mid \cdot)$.
- Accept with $p_{\text{acc}} = \min\{1, \pi(\theta')/\pi(\theta) \cdot q(\theta \mid \theta')/q(\theta' \mid \theta)\}$.
- Use log-scale; enforce support constraints by immediate rejection.

4) Practice:

- Example 5.3: non-conjugate prior $\mu > 0 \Rightarrow$ MH sampling + burn-in + posterior summaries.
- Example 5.4: Pareto + gamma prior $\Rightarrow$ MH with normal random walk and a clean log acceptance.