Computational Statistics

Section 7.3: Independent Sampler and its comparison with random-walk MH

What should we understand by the end?

1. Construct an MH algorithm using a fixed proposal distribution.
2. Derive and interpret its acceptance probability.
3. Explain why a good approximation to the target helps.
4. Compare independent proposals with random-walk proposals.
5. Recognise when a global proposal improves sampling efficiency.

Throughout, π denotes a normalised target density or probability mass function, unless stated otherwise.

Recall: the Metropolis–Hastings update

Given the current state $X_t = x$:

1. Draw a candidate $Y \sim q(\cdot \mid x)$.
2. Compute
$$\alpha(x, Y) = \min\left\{1, \frac{\pi(Y)q(x \mid Y)}{\pi(x)q(Y \mid x)}\right\}.$$
3. Draw $U \sim \text{Uniform}(0, 1)$ independently and set
$$X_{t+1} = \begin{cases} Y, & U \le \alpha(x, Y), \\ x, & U > \alpha(x, Y). \end{cases}$$

One iteration always produces a state

A rejection produces another copy of the current state.

Random walk or independent proposal?

Random-walk MH

$$Y = X_t + \sigma Z, \quad Z \sim N(0,1).$$

Thus $Y \mid X_t = x \sim N(x, \sigma^2)$.

- The proposal is centred at the current state.
- Moving the current state moves the proposal.

Independent sampler

$$Y \sim q, \qquad q(y \mid x) = q(y).$$

For example, $Y \sim N(m, s^2)$ for fixed m, s.

- The proposal is fixed across iterations.
- It can propose moves far from the current state.

Both are MH algorithms. Here, "usual MH" means random-walk MH.

Deriving the independent-sampler acceptance rule

Because the proposal does not depend on the current state,

$$q(y \mid x) = q(y), \qquad q(x \mid y) = q(x).$$

Substituting into the MH formula gives

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

Independence does not imply symmetry

In general, $q(x) \neq q(y)$. The proposal terms do not cancel merely because the proposal is independent of the current state.

An interpretation using importance weights

Define

$$w(x) = \frac{\pi(x)}{q(x)} \qquad \Longrightarrow \qquad \boxed{\alpha(x, y) = \min\{1, w(y)/w(x)\}.}$$

- Large $w(x)$: the proposal assigns relatively little probability compared with the target.
- Small $w(x)$: the proposal assigns relatively much probability compared with the target.
- If $w(y) \geq w(x)$, accept the candidate with probability 1.
- If $w(y) < w(x)$, accept with probability $w(y)/w(x)$.

Connection with importance sampling

The same ratio π/q appears, but here it determines a transition probability rather than a weighted estimator.

The independent sampler: complete algorithm

Choose a sampleable q with $q(x) > 0$ wherever $\pi(x) > 0$. Start at X_0 with $\pi(X_0) > 0$.

1. For $t = 0, \ldots, N-1$, draw $Y \sim q$.
2. Compute $\alpha = \min\{1, w(Y)/w(X_t)\}$.
3. Independently draw $U \sim \text{Uniform}(0, 1)$.
4. If $U \leq \alpha$, set $X_{t+1} = Y$; otherwise set $X_{t+1} = X_t$.
5. Store X_{t+1} in either case.

Unknown normalising constants cancel

If $\pi(x) = \widetilde{\pi}(x)/Z$, use $\widetilde{w}(x) = \widetilde{\pi}(x)/q(x)$:

$$\frac{w(y)}{w(x)} = \frac{\widetilde{w}(y)}{\widetilde{w}(x)}.$$

Why does the algorithm preserve the target?

For $x \neq y$, the probability flow from x to y is proportional to

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

This is the **detailed-balance condition**; self-transitions satisfy it automatically.

What this proves

π is a stationary distribution. Convergence from an arbitrary initial state additionally requires the usual ergodicity conditions.

Worked example: a three-state target

Let the state space be $\{A, B, C\}$ and use a uniform proposal.

State	A	B	C
Target π	0.2	0.5	0.3
Proposal q	$1/3$	$1/3$	$1/3$
Weight $w = \pi/q$	0.6	1.5	0.9

For example,

$$\alpha(B, A) = \frac{0.6}{1.5} = 0.4, \qquad \alpha(A, B) = \min\{1, 1.5/0.6\} = 1.$$

The proposal visits all states equally often; acceptance and holding times correct this to the target proportions.

Worked example: follow the chain

Start at $X_0 = B$. Use the following candidate and uniform draws:

t	X_t	Y	$\alpha(X_t, Y)$	U	Result X_{t+1}
0	B	A	0.4	0.7	B (reject)
1	B	C	0.6	0.5	C (accept)
2	C	A	2/3	0.8	C (reject)
3	C	B	1	0.9	B (accept)

$$\underbrace{B}_{X_0}, \quad B, \quad C, \quad C, \quad B.$$

Keep the repeats

The output states depend on one another, even though the candidates are drawn independently from q.

When the proposal matches the target

If $q = \pi$, then

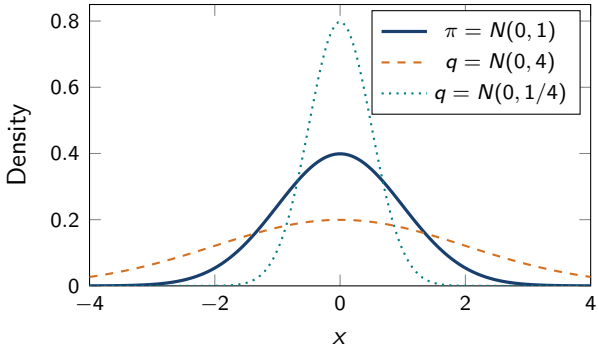
$$w(x) = 1 \quad \text{and} \quad \alpha(x, y) = 1.$$

Every candidate is accepted, and $X_1, X_2, \ldots$ are i.i.d. draws from π. If q approximates π well throughout the relevant regions, the weights vary little and acceptance is usually high.

A precise statement

An acceptance probability *close to* 1 means that most proposals are accepted. It does not mean that all proposals are accepted.

Choosing a proposal: shape and tail coverage



- A broad proposal covers the tails but wastes some proposals.
- A narrow proposal can seriously underestimate tail probability.
- Merely having positive density everywhere does not guarantee efficient sampling.

Why can an independent sampler get stuck?

For a $N(0,1)$ target and a $N(0,s^2)$ proposal,

$$w(x) = s \exp\left\{ \frac{x^2}{2} \left(\frac{1}{s^2} - 1 \right) \right\}.$$

When $s < 1$, the weights become very large in the tails. For example, let $s = 1/2$, current state $x = 2$, and candidate $y = 0$:

$$w(2) = \tfrac{1}{2} e^6, \qquad w(0) = \tfrac{1}{2}, \qquad \alpha(2,0) = e^{-6} \approx 0.0025.$$

The consequence

Once the chain reaches an underrepresented tail region, ordinary proposals near zero are often rejected. The chain can remain at the same state for many iterations.

How should we interpret a high acceptance rate?

Random-walk MH

Tiny steps can give very high acceptance:

$$Y = X_t + 0.001Z.$$

But movement is slow and successive states can be highly correlated.

Independent sampler

A well-matched q gives high acceptance while proposing fresh locations across the target distribution. This often produces more effective exploration.

Acceptance rate alone is not a convergence diagnostic

A short run can miss a poorly covered mode or tail. Inspect exploration and dependence as well as acceptance.

Comparing the two acceptance rules

Symmetric random-walk MH

$$Y = x + \sigma Z, \quad Z \sim N(0,1).$$

Because $q_{\mathrm{RW}}(y \mid x) = q_{\mathrm{RW}}(x \mid y)$,

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

Compare the target densities.

Independent MH

$$Y \sim q, \quad \text{independently of } x.$$

The reverse and forward proposals are $q(x)$ and $q(y)$, so

$$\boxed{\alpha_{\mathrm{IMH}}(x,y) = \min\{1, w(y)/w(x)\}.}$$

Compare target-to-proposal weights.

The simple random-walk formula above requires a symmetric increment distribution.

Same current state and candidate: different acceptance

Let $\pi = N(0,1)$, current state $x = 0$, and candidate $y = 2$.

Symmetric Gaussian random walk

$$\alpha_{\mathrm{RW}}(0,2) = \frac{\pi(2)}{\pi(0)} = e^{-2} \approx 0.135.$$

Independent proposal $q = N(0,4)$

Here $w(x) = 2\exp(-3x^2/8)$, giving

$$\alpha_{\mathrm{IMH}}(0,2) = \frac{w(2)}{w(0)} = e^{-3/2} \approx 0.223.$$

For $U = 0.20$, the random walk rejects and independent MH accepts.

This compares one proposed move; overall efficiency also depends on how frequently useful candidates are proposed.

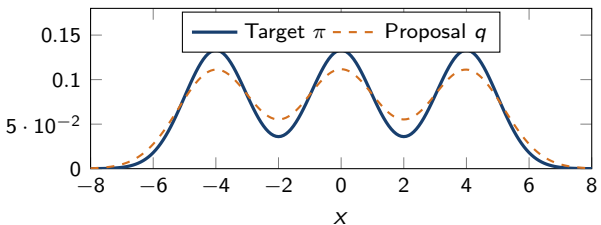
A three-mode target: constructing a global proposal

Recall the mixture from our MH example:

$$\pi(x) = \tfrac{1}{3}\phi_1(x+4) + \tfrac{1}{3}\phi_1(x) + \tfrac{1}{3}\phi_1(x-4),$$

where ϕ_s is the $N(0, s^2)$ density. Choose

$$q(x) = \tfrac{1}{3}\phi_{1.2}(x+4) + \tfrac{1}{3}\phi_{1.2}(x) + \tfrac{1}{3}\phi_{1.2}(x-4).$$



To sample from q:

1. Choose $m \in \{-4, 0, 4\}$ uniformly.
2. Draw $Y = m + 1.2Z$, with $Z \sim N(0,1)$.

The wider components cover each mode and its tails.

Benefit 1: proposing moves between distant modes

Suppose $X_t = -4$ and consider a candidate at $y = 4$.

- Both π and the mixture proposal q are symmetric, so

$$\pi(-4) = \pi(4), \qquad q(-4) = q(4).$$

- Hence **both** symmetric random-walk MH and independent MH accept this candidate with probability 1.
- But a random walk with $\sigma = 0.5$ needs a 16-standard-deviation increment to reach 4 from -4. Proposals near 4 are extremely unlikely.
- The independent mixture chooses the component centred at 4 with probability $1/3$, regardless of the current state.

The advantage is in proposing a useful move

A well-designed global proposal can jump between modes directly. Small-step random walks can spend many iterations within one mode.

Benefit 2: less dependence can improve estimation

For a quantity of interest,

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

For a stationary chain with finite variance,

$$\mathrm{Var}(\widehat{\mathcal{I}}_N) = \frac{\mathrm{Var}_\pi(f)}{N}\left[1 + 2\sum_{k=1}^{N-1}\left(1 - \frac{k}{N}\right)\rho_k\right],$$

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

- Small random-walk steps often lead to persistent positive correlations.
- A good independent proposal can reduce these correlations and improve precision for the same number of iterations.
- Compare efficiency per unit of computing time: constructing and sampling q also has a cost.

Random-walk MH versus independent MH

	Symmetric random-walk MH	Independent MH
Candidate	$Y = X_t + \sigma Z$	$Y \sim q$
Acceptance	$\min\{1, \pi(Y)/\pi(X_t)\}$	$\min\{1, w(Y)/w(X_t)\}$
Design choice	Step size and increment shape	Global proposal shape and coverage
Main strength	Simple local proposals	Distant moves when q is good
Main difficulty	Slow exploration or many rejections	Poor mode or tail coverage
Target knowledge	No global approximation needed	Good global approximation helps
After rejection	Record X_t again	Record X_t again

When is the independent sampler worth using?

A useful global approximation is already available

Use a tractable density q that can be sampled and evaluated, for example a Gaussian approximation or a fitted mixture covering the important modes. Direct samples from q generally have the wrong target distribution. The MH acceptance step corrects the transition so that π is stationary.

- Independent MH can benefit from knowledge of the target's overall shape.
- Random-walk MH is a useful starting point when a global approximation is difficult to construct.
- Both methods allow an unknown target normalising constant, retain rejected states, and require convergence checks.

No universal winner

A poor independent proposal can perform worse than a well-tuned random walk.

What should we remember?

1. The independent sampler and the random walk are both MH algorithms.
2. Independent MH uses a fixed proposal and compares two weights:

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

3. A good global proposal can improve movement between modes and reduce sample dependence.
4. These benefits depend on covering the target's important regions and tails.
5. High acceptance alone does not establish efficient exploration; repeated states must be retained in both algorithms.

Check your understanding

1. Why do the proposal terms cancel for a symmetric random walk, but not for a general independent proposal?
2. For $\pi = N(0, 1)$ and a move from 0 to 2, what is the symmetric random-walk acceptance probability?
3. For the same move with independent $q = N(0, 4)$, what is the acceptance probability?
4. Why can independent MH move between distant modes more easily, even if both methods would accept the same move?
5. Is independent MH always more efficient? What happens after a rejection in either method?

Check your understanding: answers

1. Symmetry gives $q_{\mathrm{RW}}(y \mid x) = q_{\mathrm{RW}}(x \mid y)$. Independence gives $q(y)$ and $q(x)$, which need not be equal.
2. Random walk: $\alpha_{\mathrm{RW}}(0, 2) = e^{-2} \approx 0.135$.
3. Independent proposal: $\alpha_{\mathrm{IMH}}(0, 2) = e^{-3/2} \approx 0.223$.
4. A suitable global proposal generates candidates near other modes frequently; a small-step random walk rarely proposes such a large move.
5. No: performance depends on the proposal and computing cost. Both methods record the current state again after rejection.