The Metropolis–Hastings Algorithm

1 Introduction

Suppose we wish to draw samples from a probability distribution π defined on a state space $\mathcal{X} \subseteq \mathbb{R}^d$ for which direct sampling is impractical. Typically π is known only up to a normalising constant, i.e.

$$\pi(x) = \frac{1}{Z}\, p(x), \qquad x \in \mathcal{X},\ Z = \int_{\mathcal{X}} p(x)\, \mathrm{d}x,$$

where the non–negative function p can be evaluated point-wise, but the integral Z is intractable.

The *Metropolis–Hastings* (MH) algorithm constructs a Markov chain whose stationary distribution is π, yielding asymptotically exact samples via simulation.

2 Algorithm

Let $q(y \mid x)$ be a *proposal (transition) kernel*, i.e. a conditional probability density (or mass) function from which we can sample efficiently. Common choices include symmetric random walks and independence proposals.

Algorithm 1 Metropolis–Hastings

Require: unnormalised density p, proposal kernel q, initial state x_1

for $t = 1, 2, \ldots$ **do**
    Sample $y \sim q(\cdot \mid x_t)$ ▷ *proposal step*
    Compute the acceptance probability

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

    Draw $u \sim \mathrm{Uniform}(0, 1)$
    if $u < a$ **then**
        $x_{t+1} \leftarrow y$ ▷ *accept*
    else
        $x_{t+1} \leftarrow x_t$ ▷ *reject*
    end if
end for

2.1 Examples of proposals

Gaussian random walk. $q(y \mid x) = \mathcal{N}(x, \sigma^2 I)$, where σ controls the step size. This proposal is symmetric, i.e. $q(y \mid x) = q(x \mid y)$. In this case, the acceptance probability is simply

$$a = \min\left(1, \frac{p(y)}{p(x_t)}\right).$$

This is the original Metropolis algorithm, which is a special case of MH.

Independence proposal. $q(y \mid x) = q(y)$, independent of x. The acceptance probability becomes

$$a = \min\left(1, \frac{p(y)\, q(x_t)}{p(x_t)\, q(y)}\right).$$

Independence proposals can be efficient if q is a good approximation to π, but may lead to low acceptance rates if q is poorly chosen. Independence proposals are also not symmetric.

3 Convergence

3.1 Detailed balance

Algorithm 1 defines a Markov chain with transition kernel $K(x, dy)$ given by

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

where

$$A(x, y) = \min\left(1, \frac{p(y)\, q(x \mid y)}{p(x)\, q(y \mid x)}\right), \qquad r(x) = 1 - \int_{\mathcal{X}} q(z \mid x)\, A(x, z)\, dz,$$

and δ_x denotes the Dirac measure at x. The first term describes proposed moves that are accepted, while the second term corresponds to rejected moves, ensuring that $K(x, \cdot)$ is a probability measure.

Proposition 1 (Detailed balance)**.** *The MH transition kernel K satisfies the detailed balance condition with respect to π, i.e.*

$$\pi(dx)\, K(x, dy) = \pi(dy)\, K(y, dx) \quad \text{on } \mathcal{X} \times \mathcal{X}.$$

Proof. We verify the equality separately on the off–diagonal set $\{(x, y) : x \neq y\}$ and on the diagonal.

Off–diagonal part. For $x \neq y$, we have

$$K(x, dy) = q(y \mid x)\, A(x, y)\, dy,$$

so that

$$\pi(dx) K(x, dy) = \pi(x)\, q(y \mid x)\, A(x, y)\, dx\, dy.$$

Using the definition of A and the identity $\min(1, r) = r\, \min(1, 1/r)$, we obtain

$$\begin{aligned} \pi(x)\, q(y \mid x)\, A(x, y) &= \pi(x)\, q(y \mid x)\, \frac{p(y)\, q(x \mid y)}{p(x)\, q(y \mid x)}\, A(y, x) \\ &= \frac{p(y)}{Z}\, q(x \mid y)\, A(y, x) \\ &= \pi(y)\, q(x \mid y)\, A(y, x), \end{aligned}$$

which shows that

$$\pi(dx) K(x, dy) = \pi(dy) K(y, dx) \qquad \text{for } x \neq y.$$

Diagonal part. On the diagonal, the kernel is given by the rejection mass

$$K(x, dy) = r(x)\, \delta_x(dy),$$

so that

$$\pi(dx) K(x, dy) = \pi(dx)\, r(x)\, \delta_x(dy).$$

By symmetry, we also have

$$\pi(dy) K(y, dx) = \pi(dy)\, r(y)\, \delta_y(dx).$$

Both measures are supported on the diagonal $\{(x, x) : x \in \mathcal{X}\}$ and, for any measurable set $B \subset \mathcal{X} \times \mathcal{X}$,

$$\int \mathbf{1}_B(x, y)\, \pi(dx)\, r(x)\, \delta_x(dy) = \int \mathbf{1}_B(x, x)\, \pi(dx)\, r(x),$$

which is identical to the corresponding expression with x replaced by y. Hence the two measures coincide on the diagonal.

Combining the off–diagonal and diagonal parts yields

$$\pi(dx)\,K(x,dy) = \pi(dy)\,K(y,dx),$$

completing the proof. □

3.2 Stationarity and Convergence

The detailed balance condition implies that π is *invariant* for the Markov kernel K, i.e.

$$\int_{\mathcal{X}} \pi(dx)\,K(x,dy) = \pi(dy).$$

Equivalently, if $X_n \sim \pi$, then $X_{n+1} \sim \pi$.

To obtain stronger properties such as uniqueness and convergence, additional assumptions on the chain are required. In general state spaces, the appropriate notion of irreducibility is π-irreducibility, and recurrence must be understood in the sense of Harris.

Assume that the chain is π-irreducible and Harris recurrent. Then π is the *unique* invariant probability measure.

A sufficient condition for π-irreducibility is that the proposal kernel satisfies

$$q(y\,|\,x) > 0 \quad \text{for all } x, y \in \operatorname{supp}(\pi),$$

so that every set of positive π-measure can be reached with positive probability. More generally, it suffices that for any measurable set $A \subset \mathcal{X}$ with $\pi(A) > 0$ and any $x \in \mathcal{X}$, there exists $n \geq 1$ such that $K^n(x, A) > 0$.

If, in addition, the chain is aperiodic, then for any initial distribution, the law of X_n converges to π in total variation.

Aperiodicity typically follows from the presence of self-transitions. In the Metropolis–Hastings algorithm,

$$K(x, \{x\}) = 1 - \int_{\mathcal{X}} q(y\,|\,x)\,A(x,y)\,dy,$$

which is strictly positive whenever proposals are rejected with positive probability. In that case, the chain cannot be trapped in deterministic cycles.

Under the above conditions, the ergodic theorem holds: for any measurable function $h : \mathcal{X} \to \mathbb{R}$ with $h \in L^1(\pi)$,

$$\frac{1}{T}\sum_{t=1}^{T} h(X_t) \xrightarrow{T\to\infty} \int_{\mathcal{X}} h(x)\,\pi(dx) \quad \text{almost surely.}$$

Thus, empirical averages along the MH trajectory provide consistent estimators of expectations under π.