

STAT 238 - Bayesian Statistics

Lecture Thirty Two

Spring 2026, UC Berkeley

Aditya Guntuboyina

17 April 2026

We are discussing standard MCMC techniques to obtain samples from a probability π . All these techniques are based on the Metropolis-Hastings scheme, where, given the current value x of the chain, a proposal y is first generated according to some transition kernel $Q(x, y)$, and then the proposal is accepted with probability:

$$\min\left(1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}\right). \quad (1)$$

Different MCMC methods differ on how the proposal kernel $Q(x, y)$ is chosen. We shall look at three main techniques for choosing $Q(x, y)$: Random Walk Metropolis (RWM), Metropolis Adjusted Langevin Algorithm (MALA) and Hamiltonian Monte Carlo (HMC).

1 RWM

Here y is generated from x via:

$$y = x + \sigma z \quad \text{where } z \sim N(0, I_d).$$

In other words, the proposal kernel $Q(x, y)$ is given by:

$$Q(x, y) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^d \exp\left(-\frac{\|y - x\|^2}{2\sigma^2}\right).$$

Clearly this proposal is symmetric in the sense that $Q(x, y) = Q(y, x)$. As a result the acceptance probability (1) becomes simply

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

When the dimension d is small, this method works reasonably well. When d is large, it will work only if σ is chosen very small. However, with such small σ , one would need to run the chain for a very large amount of time to ensure adequate mixing. The consequence is that RWM does not work well when d is large.

1.1 Example: π is $N(0, I_d)$

As an illustrative example, consider the case where:

$$\pi(x) \propto \exp(-\|x\|^2/2).$$

In this case, we have

$$\frac{\pi(y)}{\pi(x)} = \exp\left[\frac{1}{2}(\|x\|^2 - \|y\|^2)\right].$$

Because the proposed value y will be $x + \sigma z$ for $z \sim N(0, I_d)$, we have

$$\frac{\pi(y)}{\pi(x)} = \exp\left[\frac{1}{2}(\|x\|^2 - \|x + \sigma z\|^2)\right] = \exp(-A)$$

where

$$A := \sigma \langle z, x \rangle + \frac{\sigma^2}{2} \|z\|^2.$$

Because $z \sim N(0, I_d)$, it is easy to check that the mean and variance of A satisfy:

$$\mathbb{E}A = \frac{d\sigma^2}{2}$$

and

$$\begin{aligned} \text{var}(A) &= \mathbb{E}\left(\sigma \langle z, x \rangle + \frac{\sigma^2}{2}(\|z\|^2 - d)\right)^2 \\ &\leq 2\sigma^2 \text{var}(\langle z, x \rangle) + \frac{\sigma^4}{2} \text{var}(\|z\|^2) \leq 2\sigma^2 \|x\|^2 + 2d\sigma^4 \end{aligned}$$

which means that the standard deviation of A is $O(\sigma\|x\| + \sigma^2\sqrt{d})$. Suppose now that $\|x\| = O(\sqrt{d})$ and $\sigma \leq 1$, then the standard deviation of A is:

$$\text{std}(A) = O(\sigma\|x\| + \sigma^2\sqrt{d}) = O(\sigma\sqrt{d} + \sigma^2\sqrt{d}) = O(\sigma\sqrt{d}).$$

We thus have:

$$\mathbb{E}A = \frac{d\sigma^2}{2} \quad \text{and} \quad \text{std}(A) = O(\sigma\sqrt{d}).$$

We can now consider two regimes of σ .

1. **Case 1:** $\sigma \gg d^{-1/2}$. Then $\sigma\sqrt{d} \rightarrow \infty$ so that $\mathbb{E}A \gg \text{std}(A)$. Then (by standard concentration), A is tightly concentrated around its mean $d\sigma^2/2 \rightarrow \infty$. Therefore the acceptance probability satisfies

$$\alpha(x, y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right) = \min(1, e^{-A}) \approx e^{-d\sigma^2/2} \rightarrow 0,$$

which means that proposals are almost always rejected and the chain cannot mix.

2. **Case 2:** $\sigma = O(d^{-1/2})$. Writing $\sigma = c/\sqrt{d}$ for a constant $c > 0$, we get

$$\mathbb{E}A = \frac{c^2}{2} = O(1), \quad \text{std}(A) = O(c) = O(1).$$

Now A is an $O(1)$ random variable fluctuating around an $O(1)$ mean so e^{-A} is a positive $O(1)$ quantity. The acceptance probability is then bounded away from zero so that chain can actually move.

The conclusion is that the acceptance probability of RWM is non-degenerate if and only if $\sigma = O(d^{-1/2})$. Equivalently, the proposal step must shrink as the dimension grows, with $\sigma \sim c/\sqrt{d}$ being the critical scaling. This is the fundamental tuning constraint for RWM in high-dimensions. When d is large, $O(d^{-1/2})$ tends to be too small with the consequence that the chain takes a very large amount of samples to sample from the whole of π . This makes RWM very inefficient in high-dimensions.

2 MALA

Here y is generated from x via:

$$y = x + \frac{\sigma^2}{2} \nabla \log \pi(x) + \sigma z \quad \text{where } z \sim N(0, I_d). \quad (2)$$

In other words, the proposal kernel $Q(x, y)$ is given by:

$$Q(x, y) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^d \exp \left(-\frac{\|y - x - (\sigma^2/2)\nabla \log \pi(x)\|^2}{2\sigma^2} \right).$$

Clearly this proposal is non-symmetric in the sense that $Q(x, y) \neq Q(y, x)$ so the acceptance probability is now:

$$\min \left(1, \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} \right).$$

As in RWM, it is important to tune σ appropriately. When d is large, one still needs to choose σ small. However, σ does not have to be as small as in the case of RWM for MALA to work well. This can be easily illustrated in the $N(0, I_d)$ example.

2.1 Example: $\pi = N(0, I_d)$

Here $\pi(x) \propto \exp(-\|x\|^2/2)$ so that $\nabla \log \pi(x) = -x$. Then (below $a = 1 - (\sigma^2/2)$):

$$\begin{aligned} \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} &= \exp \left(\frac{\|x\|^2 - \|y\|^2}{2} \right) \exp \left(\frac{\|y - ax\|^2 - \|x - ay\|^2}{2\sigma^2} \right) \\ &= \exp \left(\frac{\|x\|^2 - \|y\|^2}{2} \right) \exp \left(\frac{1 - a^2}{2\sigma^2} (\|y\|^2 - \|x\|^2) \right) \\ &= \exp \left(\frac{\sigma^2}{8} (\|x\|^2 - \|y\|^2) \right). \end{aligned}$$

Plugging in

$$y = x + \frac{1}{2}\sigma^2 \nabla \log \pi(x) + \sigma z = ax + \sigma z,$$

we obtain

$$\begin{aligned} \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} &= \exp \left(\frac{\sigma^2}{8} (\|x\|^2 - \|ax + \sigma z\|^2) \right) \\ &= \exp \left(\frac{\sigma^2}{8} [\sigma^2(1 - (\sigma^2/4))\|x\|^2 - \sigma^2\|z\|^2 - 2a\sigma \langle x, z \rangle] \right) = \exp(-B) \end{aligned}$$

where

$$B := \frac{\sigma^2}{8} [\sigma^2 \|z\|^2 + 2a\sigma \langle x, z \rangle - \sigma^2(1 - (\sigma^2/4))\|x\|^2].$$

Below we use C for a generic positive constant whose value might change from specific occurrence to occurrence. Clearly

$$\mathbb{E}B = \frac{\sigma^4}{8} [d - (1 - (\sigma^2/4))\|x\|^2] = \frac{\sigma^4}{8} \left((d - \|x\|^2) + \frac{\sigma^2}{4}\|x\|^2 \right).$$

Also

$$\text{var}(B) \leq \frac{\sigma^4}{32} (\text{var}(\sigma^2\|z\|^2) + \text{var}(2a\sigma \langle x, z \rangle)) = \frac{\sigma^4}{32} (2\sigma^4 d + 4a^2\sigma^2\|x\|^2).$$

Because $a = 1 - \sigma^2/2 \leq 1$, we can write

$$\text{var}(B) \leq C\sigma^6 (d\sigma^2 + \|x\|^2)$$

so that

$$\text{std}(B) \leq C\sigma^3(\sigma\sqrt{d} + \|x\|).$$

Suppose now that x is such that:

$$d - d^{2/3} \leq \|x\|^2 \leq Cd \tag{3}$$

for some constant C . Now let

$$\sigma = \ell d^{-1/6} \tag{4}$$

for a constant ℓ . Note that this value of σ is much larger than $d^{-1/2}$ so that RWM proposals will always be rejected with this σ . But for the random variable B , we have

$$\mathbb{E}B = \frac{\sigma^4}{8} \left((d - \|x\|^2) + \frac{\sigma^2}{4}\|x\|^2 \right) \leq \frac{\sigma^4}{8} \left(d^{2/3} + \frac{\sigma^2}{4}Cd \right) \leq \frac{\ell^4}{8} + \frac{C\ell^6}{32}$$

which is a constant. Also

$$\text{std}(B) \leq C\sigma^4\sqrt{d} + C\sigma^3\|x\| \leq C\ell^4 d^{-2/3}\sqrt{d} + C\ell^3 C$$

which is also bounded by a constant. Thus for (4) (and under the assumption (3) on x), the random variable B will be bounded above by a constant with high probability, which means that the acceptance probability will remain constant. This is in sharp contrast to the case of RWM where the acceptance probability will be extremely small for (4). This argument indicates that MALA will be more effective than RWM when d is large. Specifically, MALA can be expected to work for much higher values of σ which will lead to faster mixing.

3 HMC

In HMC, the proposal y is generated by solving the following ODE:

$$x(0) = x \quad \dot{x}(0) = z \quad \ddot{x}(t) = \nabla \log \pi(x(t)) \quad y = x(\sigma). \tag{5}$$

Note that if $\nabla \log \pi(x(t))$ is replaced by $\nabla \log \pi(x)$, then it is easy to check that y equals the MALA proposal (2).

The ODE (5) is, in most cases, impossible to solve exactly. But in the Gaussian case, it is easy to write a formula for $y = x(\sigma)$. This is illustrated next.

3.1 $\pi = N(0, I_d)$

Here the ODE (5) becomes

$$x(0) = x \quad \dot{x}(0) = z \quad \ddot{x}(t) = -x(t) \quad y = x(\sigma).$$

The general solution to $\ddot{x}(t) = -x(t)$ is given by $\alpha \cos t + \beta \sin t$. The initial conditions $x(0) = x$ and $\dot{x}(0) = z$ then give

$$x(t) = x \cos t + z \sin t$$

which implies that

$$y = x(\sigma) = x \cos \sigma + z \sin \sigma.$$

The proposal transition kernel $Q(x, y)$ is therefore

$$Q(x, y) = \left(\frac{1}{\sqrt{2\pi} |\sin \sigma|} \right)^d \exp \left(-\frac{\|y - x \cos \sigma\|^2}{2 \sin^2 \sigma} \right).$$

Note now that

$$\begin{aligned} \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} &= \exp \left(\frac{1}{2}(\|x\|^2 - \|y\|^2) \right) \exp \left(\frac{\|y - x \cos \sigma\|^2 - \|x - y \cos \sigma\|^2}{2 \sin^2 \sigma} \right) \\ &= \exp \left(\frac{1}{2}(\|x\|^2 - \|y\|^2) \right) \exp \left(\frac{1}{2}(\|y\|^2 - \|x\|^2) \right) = 1. \end{aligned}$$

So the acceptance probability of this chain equals 1 **regardless of the value of σ** . This means that the HMC chain can be evolved to a large value of σ without affecting the acceptance probability at all. This is quite unlike the case for RWM and MALA where σ has to be small otherwise the acceptance probability will be very small. The possibility of taking large steps (in terms of σ) is one of the main advantages of HMC.

We shall, in the next lecture, that, for general π (not just Gaussian), the HMC proposal (5) satisfies detailed balance with respect to π for every σ (so that the subsequent M-H acceptance probability equals 1). In practice, however, the HMC proposal (5) cannot be directly used because this ODE cannot be solved exactly. One uses a numerical approximate solution to the ODE (5). This approximation spoils the detailed balance property however, so subsequent Metropolization is necessary. However, the acceptance probability (while not exactly equal to 1) will still be usually high for HMC even for not too small σ . More details will be provided next week.