MCMC Methods: Gibbs Sampling and the Metropolis-Hastings Algorithm

Patrick Lam

Metropolis-Hastings Algorithm

The Metropolis-Hastings Algorithm follows the following steps:

1. Choose a starting value $\boldsymbol{\theta}^{(0)}$.
2. At iteration t, draw a candidate $\boldsymbol{\theta}^*$ from a jumping distribution $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$.
3. Compute an acceptance ratio (probability):

$$r = \frac{p(\boldsymbol{\theta}^*|\mathbf{y})/J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})}{p(\boldsymbol{\theta}^{(t-1)}|\mathbf{y})/J_t(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*)}$$

4. Accept $\boldsymbol{\theta}^*$ as $\boldsymbol{\theta}^{(t)}$ with probability $\min(r, 1)$. If $\boldsymbol{\theta}^*$ is not accepted, then $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$.
5. Repeat steps 2-4 M times to get M draws from $p(\boldsymbol{\theta}|\mathbf{y})$, with optional burn-in and/or thinning.

Step 1: Choose a starting value $\boldsymbol{\theta}^{(0)}$.

This is equivalent to drawing from our initial stationary distribution.

The important thing to remember is that $\boldsymbol{\theta}^{(0)}$ must have positive probability.

$$p(\boldsymbol{\theta}^{(0)}|\mathbf{y}) > 0$$

Otherwise, we are starting with a value that cannot be drawn.

Step 2: Draw $\boldsymbol{\theta}^*$ from $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$.

The jumping distribution $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$ determines where we move to in the next iteration of the Markov chain (analogous to the transition kernel). The support of the jumping distribution must contain the support of the posterior.

The original **Metropolis algorithm** required that $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$ be a symmetric distribution (such as the normal distribution), that is

$$J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)}) = J_t(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*)$$

We now know with the Metropolis-Hastings algorithm that symmetry is unnecessary.

If we have a symmetric jumping distribution that is dependent on $\boldsymbol{\theta}^{(t-1)}$, then we have what is known as **random walk Metropolis sampling**.

If our jumping distribution does not depend on $\boldsymbol{\theta}^{(t-1)}$,

$$J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)}) = J_t(\boldsymbol{\theta}^*)$$

then we have what is known as **independent Metropolis-Hastings sampling**.

Basically all our candidate draws $\boldsymbol{\theta}^*$ are drawn from the same distribution, regardless of where the previous draw was.

This can be extremely efficient or extremely inefficient, depending on how close the jumping distribution is to the posterior.

Generally speaking, chain will behave well only if the jumping distribution has heavier tails than the posterior.

Step 3: Compute acceptance ratio r.

$$r = \frac{p(\boldsymbol{\theta}^*|\mathbf{y})/J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})}{p(\boldsymbol{\theta}^{(t-1)}|\mathbf{y})/J_t(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*)}$$

In the case where our jumping distribution is symmetric,

$$r = \frac{p(\boldsymbol{\theta}^*|\mathbf{y})}{p(\boldsymbol{\theta}^{(t-1)}|\mathbf{y})}$$

If our candidate draw has higher probability than our current draw, then our candidate is better so we definitely accept it. Otherwise, our candidate is accepted according to the ratio of the probabilities of the candidate and current draws.

Note that since r is a ratio, we only need $p(\boldsymbol{\theta}|\mathbf{y})$ *up to a constant of proportionality* since $p(\mathbf{y})$ cancels out in both the numerator and denominator.

In the case where our jumping distribution is not symmetric,

$$r = \frac{p(\boldsymbol{\theta}^*|\mathbf{y})/J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})}{p(\boldsymbol{\theta}^{(t-1)}|\mathbf{y})/J_t(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*)}$$

We need to weight our evaluations of the draws at the posterior densities by how likely we are to draw each draw.

For example, if we are very likely to jump to some $\boldsymbol{\theta}^*$, then $J_t(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})$ is likely to be high, so we should accept less of them than some other $\boldsymbol{\theta}^*$ that we are less likely to jump to.

In the case of independent Metropolis-Hastings sampling,

$$r = \frac{p(\boldsymbol{\theta}^*|\mathbf{y})/J_t(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(t-1)}|\mathbf{y})/J_t(\boldsymbol{\theta}^{(t-1)})}$$

Step 4: Decide whether to accept $\boldsymbol{\theta}^*$.

Accept $\boldsymbol{\theta}^*$ as $\boldsymbol{\theta}^{(t)}$ with probability $\min(r, 1)$. If $\boldsymbol{\theta}^*$ is not accepted, then $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$.

1. For each $\boldsymbol{\theta}^*$, draw a value u from the Uniform(0,1) distribution.
2. If $u \leq r$, accept $\boldsymbol{\theta}^*$ as $\boldsymbol{\theta}^{(t)}$. Otherwise, use $\boldsymbol{\theta}^{(t-1)}$ as $\boldsymbol{\theta}^{(t)}$

Candidate draws with higher density than the current draw are always accepted.

Unlike in rejection sampling, each iteration always produces a draw, either $\boldsymbol{\theta}^*$ or $\boldsymbol{\theta}^{(t-1)}$.

Acceptance Rates

It is important to monitor the *acceptance rate* (the fraction of candidate draws that are accepted) of your Metropolis-Hastings algorithm.

If your acceptance rate is too high, the chain is probably not mixing well (not moving around the parameter space quickly enough).

If your acceptance rate is too low, your algorithm is too inefficient (rejecting too many candidate draws).

What is too high and too low depends on your specific algorithm, but generally

- random walk: somewhere between 0.25 and 0.50 is recommended
- independent: something close to 1 is preferred