Markov Chain Monte Carlo - 1

By Matthew Stephens (edited by Heejung Shim)

Markov chain Monte Carlo (MCMC)

MCMC is a very widely-used technique for simulating samples from the posterior distribution.

The basic idea is to simulate a Markov chain, whose stationary distribution is the posterior distribution. By simulating the Markov chain for long enough, one obtains samples that are "approximately" from the posterior distribution.

- Metropolis–Hastings Algorithm
- Gibbs Sampling

Metropolis–Hastings Algorithm

Let $q(\theta, \theta')$ denote a transition density, which is a way of generating new states θ' given the current state θ. Let π denote the target distribution, up to a constant of proportionality (e.g. to simulate from the posterior, $\pi = p(\theta)p(x|\theta)$).

Metropolis–Hastings Algorithm

Now consider the following Markov chain:

1. Start with an initial value $\theta^{(0)}$; set $t = 0$.
2. Given the current value of $\theta^{(t)} = \theta$, generate a proposed new value θ' according to $q(\theta, \cdot)$.
3. Define the acceptance probability A by

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

4. With probability A set $\theta^{(t+1)} = \theta'$; otherwise set $\theta^{(t+1)} = \theta$.
5. Increase t; return to step 2.

Note: q is referred to as the *proposal distribution*. The probability A is the *acceptance probability*.

Why does Metropolis–Hastings algorithm work?

See MarkovProcesses.pdf for definition and detailed explanation.

- The process $\theta^{(t)}$ has stationary distribution π.
  - Proof: check detailed balance condition, $\pi(\theta)p(\theta \to \theta') = \pi(\theta')p(\theta' \to \theta)$.
- π will be limiting.
  - The process $\theta^{(t)}$ can stay where it is, it is aperiodic.
  - For all sensible choices of q it will also be irreducible (i.e., regardless the present state, it can reach any other state in finite time).
  - If a Markov process is aperiodic and irreducible then any stationary distribution is unique and limiting.
- The process is ergodic.
  - The process $\theta^{(t)}$ is aperiodic and irreducible, and π is a limiting distribution, then the process is ergodic.

Why does Metropolis–Hastings algorithm work?

- Because the output process $\theta^{(t)}$ has a limiting distribution, the output of the M-H algorithm will—in the limit—be a sequence of observations from the density π.
- These observations are not independent, but because they come from an ergodic process we can still use them to estimate properties of π.

Metropolis–Hastings Algorithm

Remarks

1. For this algorithm to work q can be quite general, but we do need to know how to simulate from $q(\theta, \cdot)$ for every θ.
2. The algorithm only requires that we know π up to constant.

Choosing the proposal distribution q

Ideally you want q to have two properties

- It proposes large moves.
- It leads to high average acceptance probabilities, A.

In general these two properties are competing: it is difficult to achieve both. As a result, one generally tries to choose q so that it gives "intermediate" average values for A [An exception to this is Gibbs sampling, where $A = 1$.]

Choosing the proposal distribution q

Example: Random Walk Metropolis-Hastings Perhaps the

most common type of MH sampler is the "random walk" MH sampler, where the proposal distribution q involves adding a symmetric random number (e.g. $U[-\epsilon, \epsilon]$ or $N(0, \epsilon^2)$) to the current value of θ.

In this case, the terms involving q in the acceptance probability cancel, due to symmetry. That is $q(\theta, \theta') = q(\theta', \theta)$ and

$$A = \min\left(1, \frac{\pi(\theta')}{\pi(\theta)}\right).$$

The value of ϵ determines the typical size of the proposed move, and hence the typical value for A.

Random-walk M-H

The simpler form of A in the random-walk M-H makes it easier to see what the algorithm does.

The proposal process has no preferences regards which part of the parameter space it wanders in.
If the proposal process wants to move somewhere where π is larger, then $A = 1$ and the move is always accepted.
However if the proposal process wants to move somewhere where π is smaller, then $A < 1$ and there is a chance we stay where we are.
In this way the M-H process will spend relatively more time in regions where π is large.

Example

https://theoreticalecology.wordpress.com/2010/09/17/metropolis-hastings-mcmc-in-r/

`MHexample.pdf`