How does the Metropolis-Hastings Algorithm Work?

Acknowledgement

This exposition closely follows the logic in S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings Algorithm", *The American Statistician*, 49(4):327-335. 1995.

Definitions

Call $f(x)$ the target density for parameters x, which lie in some (discrete or continuous) parameter space. We want to build a Markov Chain that will "explore" the parameter space, and which will, in the limit, visit any two points x and y with relative frequencies $f(x)$and $f(y)$.

Call $\tau(x \to y)$ the one-step transition probability of the search process. It tells us the probability of moving in one search step from any specific point x in the parameter space to any other point y, with $\sum_y \tau(x \to y) = 1$ over the set of all possible y values (including x).

We build the transition function $\tau()$ using a reject/accept procedure with two steps: when the chain is at point x, we (1) propose a transition to y with probability $q(x \to y)$, and then (2) accept that proposal with probability $\alpha(x \to y)$. Thus

$$\tau(x \to y) = q(x \to y) \cdot \alpha(x \to y)$$

If the proposal is rejected then the chain stays at x. Note that the proposal distribution can be *any* convenient distribution.

Necessary Long-run Properties of the Transition Rule $\tau()$

We want a Markov process that will select realization x with probability $f(x)$, for every x in the parameter space. This requires that

$$f(x) = \sum_y f(y)\,\tau(y \to x)$$

where the sum is over all other possible parameters y. If this condition is satisfied, then the chain will converge to the target distribution: each parameter value is visited with a limiting frequency proportional to its probability $f()$.

The convergence condition is satisfied if

$$f(x)\,\tau(x \to y) \;=\; f(y)\,\tau(y \to x)$$

because summing both sides over all possible y values produces

$$\begin{aligned} f(x) \sum_y \tau(x \to y) &= \sum_y f(y)\,\tau(y \to x) \\ f(x) \cdot 1 &= \sum_y f(y)\,\tau(y \to x) \end{aligned}$$

Thus a transition rule with

$$f(x)\,\tau(x \to y) \;=\; f(y)\,\tau(y \to x)$$

will produce a chain that converges to $f()$.

With a reject/accept procedure, we therefore want a proposal function $q(x \to y)$ and an acceptance probability $\alpha(x \to y)$ such that

$$f(x)\,q(x \to y)\,\alpha(x \to y) \;=\; f(y)\,q(y \to x)\alpha(y \to x)$$

or

$$\frac{\alpha(x \to y)}{\alpha(y \to x)} = \frac{f(y)\, q(y \to x)}{f(x)\, q(x \to y)} = R(x \to y)$$

This tells us how many times more likely we must be to accept an $x \to y$ transition than the reverse. The ratio condition is satisfied (for any target density f and any proposal distribution q) if we accept all proposed transitions $x \to y$ for which $R(x \to y) \geq 1$, and accept a fraction $R(x \to y)$ of those for which $R(x \to y) < 1$. That is

$$\alpha(x \to y) = min[\,1\,,\; R(x \to y)]$$

Summary

- A chain converges to the target distribution $f()$ if $f(x)\,\tau(x \to y) \;=\; f(y)\,\tau(y \to x)$.
- If we're using an accept/reject Markov algorithm with $\tau(x \to y) = q(x \to y) \cdot \alpha(x \to y)$ then the convergence condition requires that acceptance ratios have the form $R(x \to y)$ derived above.
- Acceptance ratios have the form above if the acceptance probabilities are $\alpha(x \to y) = min[\,1\,,\; R(x \to y)]$.