STAT 578: Advanced Bayesian Modeling

Week 7 – Lesson 3

Metropolis and Metropolis-Hastings

Fall 2019

The Metropolis Algorithm

Goal: Formulate a Markov chain

$$\theta^0, \quad \theta^1, \quad \theta^2, \quad \ldots$$

with the posterior as its stationary (and limiting) distribution.

Our next algorithms: **Metropolis** and **Metropolis-Hastings**

Both are general purpose but require some choices and tuning.

Process:

1. Define a convenient (easily simulated) random walk on the parameter space.
2. At each iteration, simulate a step from the random walk, but take the step – *accept* it – with only a certain step-dependent probability.

   If the step is not accepted, remain at (repeat) the current location.
3. Use all accepted and repeated draws as a dependent sample from the posterior.

The probability of accepting each step must be carefully set such that the stationary (and limiting) distribution is the posterior.

The Random Walk

A random walk is a Markov chain defined by a transition distribution, or **jumping distribution**, that usually has a density (or **kernel**):

$$J_t(\theta' \mid \theta)$$

This is the density for the state θ' at time t, given the state θ at time $t-1$.

In the Metropolis algorithm, this is the **proposal density** of the **proposal distribution**.

Note: Not necessarily time-invariant

The Metropolis Algorithm

Choose starting value θ^0 and symmetric proposal density $J_t(\theta' \mid \theta) = J_t(\theta \mid \theta')$.

For $t = 1,\ 2,\ \ldots$

1. Sample **proposal** θ^* from $J_t(\cdot \mid \theta^{t-1})$
2. Compute
$$r \;=\; \frac{p(\theta^* \mid y)}{p(\theta^{t-1} \mid y)}$$
3. Set
$$\theta^t \;=\; \begin{cases} \theta^* & \text{with probability } \min(r, 1) \\ \theta^{t-1} & \text{otherwise} \end{cases}$$

Continue until sufficiently many draws are taken.

The ratio at iteration t

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

compares the posterior density value at the proposal θ^* to the value at the most recent draw θ^{t-1}. It affects whether the proposal is accepted:

- If $r > 1$, then the proposal has *higher* posterior density value, and it is always accepted.

- If $r < 1$, then the proposal has *lower* posterior density value, and it is accepted only with probability r.

Successive draws will tend to move toward and stay in regions of high posterior density, only rarely wandering to other regions.

Notice: Unlike in rejection sampling, every iteration of Metropolis produces a new draw, although it may be identical to the previous draw.

The Metropolis algorithm depends on the posterior only through

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

Notice:

- Normalizing factor $p(y)$ cancels out:

$$r = \frac{p(\theta^*)\, p(y \mid \theta^*)}{p(\theta^{t-1})\, p(y \mid \theta^{t-1})}$$

- Likelihood $p(y \mid \theta)$ needed only up to proportionality in θ.

The Metropolis algorithm clearly defines a Markov chain: The next state depends only on the current state.

The Markov chain is time-invariant only if the proposal density $J_t(\cdot \mid \cdot)$ does not depend on t.

Why is the posterior stationary for this Markov chain? See BDA3, Sec. 11.2.

Note: The proposal distribution should allow all parts of the parameter space to be potentially sampled. Otherwise, the posterior distribution might not be the limiting distribution of the Markov chain.