STAT 578: Advanced Bayesian Modeling

Week 7 – Lesson 3

Metropolis and Metropolis-Hastings

Fall 2019

Using and Extending Metropolis

Symmetric Proposals

Recall symmetry condition: $J_t(\theta' \mid \theta) \;=\; J_t(\theta \mid \theta')$

Suppose

- Parameter space is a vector space (or can be extended to one).
- $g_t(\cdot)$ is a density on this space satisfying $g_t(v) = g_t(-v)$.

Then a possible symmetric proposal is

$$J_t(\theta' \mid \theta) \;=\; g_t(\theta' - \theta).$$

Can simulate v from g_t, then let $\theta' = \theta + v$.

Multivariate Normal Proposal

A very popular form of proposal is a density of the **multivariate normal distribution** $\mathrm{N}(\theta, \Sigma)$:

$$J(\theta' \mid \theta) \quad \propto \quad \exp\left(-\tfrac{1}{2}\,(\theta' - \theta)^T\,\Sigma^{-1}\,(\theta' - \theta)\right)$$

(BDA3, Table A.1)

Note:

- Symmetric
- Positive over whole space, so can reach any location
- Easily sampled

Sampling efficiency is sometimes improved with a proposal distribution that is not symmetric.

This is allowed by a generalization: The Metropolis-Hastings algorithm ...

The Metropolis-Hastings Algorithm

Choose starting value θ^0.

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

1. Sample **proposal** θ^* from $J_t(\cdot \mid \theta^{t-1})$
2. Compute
$$r \;=\; \frac{p(\theta^* \mid y) \big/ J_t(\theta^* \mid \theta^{t-1})}{p(\theta^{t-1} \mid y) \big/ J_t(\theta^{t-1} \mid \theta^*)}$$
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.

Notice:

- Includes Metropolis as a special case
- Like Metropolis, does not require the normalizing factor
- Unlike Metropolis, could require evaluation of the proposal density

The posterior can be proven to be stationary for Metropolis-Hastings, as it was for Metropolis.

Tuning

In Metropolis or Metropolis-Hastings, the overall (long-run) fraction of iterations at which the proposal is accepted is the **acceptance rate**.

The acceptance rate can be tuned using the proposal:

- Proposals highly concentrated around the current θ lead to a high acceptance rate.
- Very wide proposals (with high variance) lead to a low acceptance rate.

Theory suggests the optimal (most efficient) acceptance rate ranges from 0.44 (one dimension) to about 0.23 (many dimensions) – see BDA3, Sec. 12.2.

The acceptance rate is usually not easy to determine analytically.

It may be estimated as the the fraction of accepted proposals in a run of the chain.

A more precise estimate is the average of the acceptance probabilities $\min(r, 1)$ in a run of the chain.

Adaptation

Since acceptance rate is not known in advance, use an *adaptive* algorithm:

- Choose a type of proposal distribution whose scale (width) can be adjusted, such as the multivariate normal.
- Run Metropolis(-Hastings), periodically adjusting the proposal scale to produce a more nearly optimal acceptance rate.

(Can also adjust the whole covariance matrix Σ of the normal proposal to more closely match covariance of draws.)

Warning: During this **adaptation** phase, the algorithm is not a Markov chain – best to discard these iterations. Once an approximately "optimal" proposal is found, keep it fixed for all samples that will ultimately be used.

Metropolis within Gibbs

Recall: Gibbs sampler requires sampling from conditional posterior of each piece θ_j of θ.

Direct sampling is easy if θ_j has a partially conjugate prior but usually not otherwise.

Idea: Instead of sampling θ_j directly, use a step of Metropolis(-Hastings).