STAT 578: Advanced Bayesian Modeling

Week 7 – Lesson 2

Gibbs Sampling

Fall 2019

The Gibbs Sampler

Goal: Formulate a Markov chain

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

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

Then simulate from this chain, eventually producing a (dependent) sample from the posterior (approximately), to use for approximating Bayesian tools.

Our first kind of Markov chain: The **Gibbs sampler**

It is the basic approach used by JAGS and OpenBUGS.

Conditional Posteriors

Suppose

$$\theta \quad = \quad (\theta_1, \ldots, \theta_d)$$

The (joint) posterior density is

$$p(\theta \mid y) \quad = \quad p(\theta_1, \ldots, \theta_d \mid y)$$

Let

$$\theta_{-j} \quad = \quad \theta \text{ without } \theta_j$$

The **conditional posterior** (or **full conditional**) distribution of θ_j has density

$$p(\theta_j \mid \theta_{-j}, y)$$

Gibbs Sampler Algorithm

Choose starting value θ^0.

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

1. Set $\theta^* =\ \theta^{t-1}$.
2. For $j = 1, \ldots, d$

   Replace θ_j^* with a sample from $p(\theta_j \mid \theta_{-j}^*, y)$.
3. Let $\theta^t =\ \theta^*$.

Continue until sufficiently many samples are taken.

In words:

Repeatedly cycle through the components of θ, updating each with a sample from its conditional posterior.

Record θ at the end of each cycle.

The Gibbs sampler is a Markov chain:

θ^t is sampled using only components of θ^{t-1}.

It is also time-invariant: The same conditional posteriors are used in each cycle.

The posterior is stationary for the Gibbs sampler:

Let θ^* have the posterior distribution, and replace θ_j^* with a draw θ_j^{**} from its conditional posterior to get θ^{**}. Then

- $\theta_{-j}^{**} = \theta_{-j}^*$, so they trivially have the same marginal posterior, and
- The conditional posterior distribution of θ_j^{**} given $\theta_{-j}^{**} = \theta_{-j}^*$ is the same as the conditional posterior distribution of θ_j^* given θ_{-j}^*.

So θ^{**} has the same distribution as θ^*.

The posterior also tends to be the limiting distribution (with rare exceptions).

Limitation: Every conditional posterior must be easy to sample.

May be true for θ_j if

- It is only one-dimensional, or
- Its conditional posterior is a standard distribution, as when the prior is chosen to make θ_j partially conjugate.

In either case, methods are available that do not require the normalizing factor.

(Later: More sophisticated sampling schemes)

A Gibbs sampler is easiest to implement when the prior is chosen to make all components of θ partially conjugate.

Even in cases where this is not feasible, *data augmentation* can sometimes restore partial conjugacy – see BDA3, Sec. 12.1.

Problem: Variates from a Gibbs sampler tend to be highly dependent when components of θ have high posterior dependence (e.g., BDA3, Fig. 11.2).

Possible remedies:

- Block updates (multivariate θ_j)
- *Parameter expansion* (BDA3, Sec. 12.1)