Johns Hopkins Engineering

625.464 Computational Statistics

Gibbs Sampling

Module 6 Lecture 6B

Why Gibbs Sampling?

multidimensional dist f

Basic idea: Build up the random vector element by element, by sequentially sampling from univariate conditional distributions

Basic Gibbs Sampler

Let $\underline{X} = (X_1, X_2, \ldots, X_p)^T$

want $\underline{X} \sim f$

Let $\underline{X}_{-i} = (X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p)^T$

$$\underline{X}_i \mid \underline{X}_{-i} \qquad f(x_i \mid \underline{X}_{-i})$$

Gibbs Sampling Algorithm

① Select starting value $\underline{X}^{(0)} = \left(X_1^{(0)}, X_2^{(0)}, \dots, X_p^{(0)}\right)^T$
and set $t=0$

② Generate in turn

$$X_1^{(t+1)} \text{ drawn from } f_1\left(X_1 \mid X_2^{(t)}, X_3^{(t)}, \dots, X_p^{(t)}\right)$$
$$X_2^{(t+1)} \text{ drawn from } f_2\left(X_2 \mid X_1^{(t+1)}, X_3^{(t)}, \dots, X_p^{(t)}\right)$$
$$\vdots$$
$$X_i^{(t+1)} \text{ drawn from } f_i\left(X_i \mid X_1^{(t+1)} \dots X_{i-1}^{(t+1)}, X_{i+1}^{(t)}, \dots X_p^{(t)}\right)$$
$$\vdots$$
$$X_p^{(t+1)} \text{ drawn from } f_p\left(X_p \mid X_1^{(t+1)} \dots X_{p-1}^{(t+1)}\right)$$

③ Increment t and go to Step ②

Comments on Gibbs Sampling

1. At each draw in Step 2, we are conditioning on the most recent update to all other elements.

2. The densities f_1, f_2, ..., f_p are called the full conditionals. Gibbs sampling only requires the full conditionals.

3. Even for high dimensional problems, all of our simulations are univariate. Advantageous, but can be slow.

4. Once convergence of the MC is achieved, the resultant vector $X^{(t)}=(X_1^{(t)}, ..., X_p^{(t)})$ is from f with each component converging individually.

5. Each draw X_i is from a marginalization of f.

Simple Gibbs Sampling Example

Want to sample $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}(\sigma, \Sigma)$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \& \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$[X_1 | X_2] = \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2),\ \sigma_1^2(1-\rho^2)\right)$$

$$[X_2 | X_1] = \mathcal{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X_1 - \mu_1),\ \sigma_2^2(1-\rho^2)\right)$$

$$X_1^{(t+1)} \sim f(X_1 | X_2^{(t)}) = \frac{e^{-1/2\frac{\left(x-\left(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(X_2^{(t)}-\mu_2)\right)\right)^2}{\sigma_1^2(1-\rho)}}}{\sqrt{2\pi\sigma_1^2(1-\rho^2)}}$$

$$X_2^{(t+1)} \sim f(X_2 | X_1^{(t+1)}) = \frac{e^{-1/2\frac{\left(x-\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(X_1^{(t+1)}-\mu_1)\right)\right)^2}{\sigma_2^2(1-\rho)}}}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}}$$

Variations and Generalizations of Gibbs Sampling

1. the order of the updates can change.

2. Blocking – updating a group of elements.

Ex/ $X_1^{(t+1)} \sim f(X_1 \mid X_2^{(t)}, X_3^{(t)}, X_4^{(t)})$

$X_2^{(t+1)}, X_3^{(t+1)} \sim f(X_2, X_3 \mid X_1^{(t+1)}, X_4^{(t)})$

$X_4^{(t+1)} \sim f(X_4 \mid X_1^{(t+1)}, X_2^{(t+1)}, X_3^{(t+1)})$

3. Hybrid Gibbs sampling
   - add metropolis hastings steps where convenient.