VI. The Gibbs Sampler

A. Introduction

To motivate the Gibbs Sampler, we consider a modification of Data Augmentation which we will refer to as Chained Data Augmentation. The Gibbs Sampler turns out to be a multivariate extension of Chained Data Augmentation.

1. Chained Data Augmentation

To begin, consider the two Data Augmentation equations:

**Predictive Equation:** 
\[ p(Z|Y) = \int_{\theta} p(Z|Y, \theta)p(\theta|Y)d\theta \]

**Posterior Equation:** 
\[ p(\theta|Y) = \int_{Z} p(\theta|Y, Z)p(Z|Y)dZ. \]

To draw an observation from the current approximation to \( p(Z|Y) \), one may use the Predictive Equation. Given the current approximation to \( p(\theta|Y) \), \( g_i(\theta) \):

a1) Generate \( \theta \) from \( g_i(\theta) \).

a2) Generate \( z \) from \( p(Z|\phi, Y) \), where \( \phi \) is the value obtained in (a1).

The Posterior Equation suggests that one may now draw a value of \( \theta \) by sampling from \( p(\theta|z, Y) \), where \( z \) is the value obtained in (a2). Given the new value for \( \theta \), step (a2) may be repeated and the algorithm iterated. It is pointed out that this iterative algorithm is Data Augmentation with \( m = 1 \). We will refer to this version of Data Augmentation as Chained Data Augmentation.

As noted in Tanner and Wong (1987), the values of \( \theta \) (over the iterations \( i = 1, 2, \ldots \)) form a Markov process with transition function equal to \( k(\theta, \phi) \), as defined in Section V.A. Under the regularity conditions of Section V.E, this is an ergodic Markov process with an equilibrium distribution satisfying the fixed point equation (***) of Section V.A. By virtue of the results in Section V.E, we have

\[ \theta^{(i)} \overset{\text{d}}{\to} \theta \sim p(\theta|Y) \]

and

\[ z^{(i)} \overset{\text{d}}{\to} Z \sim p(Z|Y), \]

where \( \theta^{(i)} \) and \( z^{(i)} \) are the sampled values at iteration \( i \).

In this manner, \( m \) iid tuples \( (\theta_j^{(i)}, z_j^{(i)}), j = 1, \ldots, m \), can be generated by repeating the path to the \( i \)th iteration \( m \) times. Clearly, the paths are independent, though there is within path dependence. Chained Data Augmentation is a natural candidate for implementation on a parallel processor machine.
Example Simulation in Hierarchical Models

Morris (1987) applied Chained Data Augmentation in the context of hierarchical models. In particular, suppose that \( d \) population means \( \theta = (\theta_1, \ldots, \theta_d) \) are to be estimated having observed \( d \) independent normally distributed sample means \( Y = (Y_1, \ldots, Y_d) \) where \( Y_i \) given \( \theta_i \) are iid \( N(\theta_i, V_i), i = 1, \ldots, d \), and the \( V_i = \text{var} (Y_i|\theta_i) \) are known. The conjugate prior distribution is taken for each \( \theta_i \) independently with \( A \) unknown and \( \theta_i \) given \( A \) are iid \( N(0, A), i = 1, \ldots, d \). The distribution on the hyperparameter \( A \) is \( cA^{-1-\eta/2}\exp(-.5\lambda/A) \), with known \( q > 0 \) and \( \lambda > 0 \).

Morris (1987) notes that these choices lead to proper posterior densities. In particular, the distribution \( p(\theta|Y, A) \) is normal, with the \( j \)th component distributed as

\[
N\{(1 - B_j)Y_j, V_j(1 - B_j)\},
\]

\( j = 1, \ldots, d \) and \( B_j = V_j/(V_j + A) \). The distribution \( p(A|\theta, Y) \) is a reciprocal chi-squared distribution

\[
\frac{\lambda + \| \theta \|^2}{\chi^2_{d+q}},
\]

where \( \| \theta \|^2 \) denotes the sum of squares. Thus, the Chained Data Augmentation algorithm is given in this case as follows. At iteration \( i \):

a1) Sample \( \theta^{(i)}_j \) from the appropriate normal, \( j = 1, \ldots, d \).

a2) Sample \( A^{(i)} \) from the appropriate chi-squared distribution.

b) Update the posterior of \( \theta_j \) as \( N\{(1 - B^{(i)}_j)Y_j, V_j(1 - B^{(i)}_j)\} \), with \( B^{(i)}_j = V_j/(V_j + A^{(i)}) \).

Morris (1987) suggests that \( m \) chained paths be created to allow for computation of posterior means and computation of posterior distributions. Note that in this case, \( A \) plays the role of the "missing" data.

2. Multivariate Chained Data Augmentation-The Gibbs Sampler

We now consider a multivariate extension of the Chained Data Augmentation algorithm. As an illustration consider three variables \( \theta, \eta \) and \( Z \). In this case, the equations are

1) \( p(Z|Y) = \int_\eta \int_\theta p(Z|Y, \theta, \eta)p(\theta, \eta|Y)d\theta d\eta \)
2) \( p(\theta|Y) = \int_\eta \int_Z p(\theta|Y, Z, \eta)p(Z, \eta|Y)dZ d\eta \)
3) \( p(\eta|Y) = \int_\theta \int_Z p(\eta|Y, Z, \theta)p(Z, \theta|Y)dZ d\theta \).

Analogous to Chained Data Augmentation these equations are used to define an iterative sampling scheme. Given starting values \( \theta^{(0)} \) and \( \eta^{(0)} \), equation 1 suggests that we sample \( z^{(1)} \sim p(Z|Y, \theta^{(0)}, \eta^{(0)}) \). Given \( z^{(1)} \) and \( \eta^{(0)} \), we sample \( \theta^{(1)} \) from \( p(\theta|Y, z^{(1)}, \eta^{(0)}) \), according to equation 2. The value \( \eta^{(1)} \) is then drawn from \( p(\eta|Y, z^{(1)}, \theta^{(1)}) \), following equation 3. The process is then iterated.