Chapter 6

The Gibbs Sampler

The Gibbs sampler is a special case of the Metropolis-Hastings algorithm. It generates a sample from an arbitrarily complex multidimensional distribution by sampling from each of the univariate full conditional distributions in turn.

Bivariate example

To keep things simple, consider sampling from an arbitrary two dimensional distribution $p(\theta_1, \theta_2)$. (In our context, this will be the posterior distribution for a Bayesian model with two parameters, $\pi(\theta_1, \theta_2 \mid y)$.)

It is assumed that the two univariate conditional densities $p(\theta_1 \mid \theta_2)$ and $p(\theta_2 \mid \theta_1)$ can be sampled from.

Bivariate example

The Gibbs sampler argument proceeds as follows –

If we are able to generate a value θ_1 from the marginal density p(θ_1) , then a value θ_2 sampled from the density p(θ_2 | θ_1) will have density function p(θ_2) and the pair of values (θ_1, θ_2) has joint density p(θ_1, θ_2).

Similarly,

If we are able to generate a value θ_2 from the marginal density p(θ_2) , then a value θ_1 sampled from the density p(θ_1 | θ_2) will have density function p(θ_1) and the pair of values (θ_1, θ_2) has joint density p(θ_1, θ_2). .

Bivariate example

If ($\theta_1^{(k)}$, $\theta_2^{(k)}$) is the sampled point at iteration k then the Gibbs sampler obtains the next point, ($\theta_1^{(k+1)}$, $\theta_2^{(k+1)}$) by generating $\theta_1^{(k+1)}$ from the density p(θ_1 | $\theta_2^{(k)}$) and $\theta_2^{(k+1)}$ from the density p(θ_2 | $\theta_1^{(k+1)}$) .

The “burn in” period is required so that the If ‘s on the previous page are satisfied.

Higher dimensional parameters

The same idea works in p (>2) dimensions, but each univariate density sampled from is a "full conditional" density, whereby all other parameters are conditioned upon.

For example, the value of $\theta_1^{(k+1)}$ is obtained by sampling from the univariate density

$$p(\theta_1 \mid \theta_2^{(k)}, \theta_3^{(k)}, \ldots, \theta_p^{(k)})$$

and then the value of $\theta_2^{(k+1)}$ is obtained by sampling from the univariate density

$$p(\theta_2 \mid \theta_1^{(k+1)}, \theta_3^{(k)}, \ldots, \theta_p^{(k)})$$

…and so on.

WinBUGS implementation

In the Bayesian context, the univariate full conditional densities that WinBUGS is working with are

$$\pi(\theta_1 \mid \theta_2^{(k)}, \theta_3^{(k)}, \ldots, \theta_p^{(k)}, y)$$

These are, for a very large class of Bayesian models, relatively straightforward to deduce. This is essentially what WinBUGS is doing during the compilation phase.

WinBUGS implementation

In particular, for models that can be drawn by doodles, the full conditional density for any stochastic node is a function of only the parent and offspring nodes. To see this, note that Bayes formula says

$$\pi(\theta_1 \mid \theta_2, \theta_3, \ldots, \theta_p, y) = \frac{\pi(\theta_1, \theta_2, \theta_3, \ldots, \theta_p, y)}{\pi(\theta_2, \theta_3, \ldots, \theta_p, y)}$$

The above full conditional density is a function of only θ_1 because all other values are treated as constants, and so it is only terms in the joint density involving θ_1 that are of relevance to the full conditional.