1 The Gibbs sampler as a special case of the Metropolis-Hastings algorithm (Based on Gelman et al BDA 2nd edn, p.293)

1.1 Preliminaries

Suppose the vector of unknown parameters is $\boldsymbol{\theta} = \theta_1, \ldots, \theta_M$. The unknowns could include missing data as well as model parameters but we will just use the generic notation $\boldsymbol{\theta}$ to represent the unknowns of interest. Our aim is to compute the posterior distribution $p(\boldsymbol{\theta}|\text{data})$.

The t^{th} iteration of a Gibbs Sampler consists of a series of steps. In each of these steps a draw is made from the full conditional distribution of each component of $\boldsymbol{\theta}$ given all other components of $\boldsymbol{\theta}$. The Gibbs sampling algorithm can be described as follows:

1 initialise $\theta_2, \theta_3, \ldots, \theta_M$ to $\theta_2^{(0)}, \theta_3^{(0)}, \ldots, \theta_M^{(0)}$

2 for (t in $1 : T$) {

2.1 draw $\theta_1^{(t)}$ from $p(\theta_1|\theta_2^{(t-1)}, \ldots, \theta_M^{(t-1)}, \text{data})$

2.2 draw $\theta_2^{(t)}$ from $p(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)} \ldots, \theta_M^{(t-1)}, \text{data})$

$\vdots$

2.M draw $\theta_M^{(t)}$ from $p(\theta_M|\theta_1^{(t)}, \theta_2^{t)} \ldots, \theta_{M-1}^{(t)}, \text{data})$ }

At convergence, the parameter values generated from this algorithm can be regarded as draws from the posterior $p(\boldsymbol{\theta}|\text{data})$.

To simplify the presentation we will use notation such as $\boldsymbol{\theta}_j$ to denote j^{th} parameter and $\boldsymbol{\theta}_{-j}$ to denote all parameters in $\boldsymbol{\theta}$ except $\boldsymbol{\theta}_j$. Thus for any j in $1, \ldots, M$ we can can write $\boldsymbol{\theta} = (\theta_j, \boldsymbol{\theta_{-j}})$.

1.2 Relating the Gibbs sampler to the Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm works with the full parameter vector, $\boldsymbol{\theta}$, and jumps through the parameter space using, at the t^{th} iteration, the jumping distribution $J_t(\boldsymbol{\theta}^{new}|\boldsymbol{\theta}^{(t-1)})$. To see the connection between the Gibbs sampler and Metropolis-Hastings algorithm we first note that, for any j in $(1 \dots M)$, we can write $\boldsymbol{\theta}^{new} = (\theta_j^{new}, \boldsymbol{\theta}_{-j}^{new})$ and $\boldsymbol{\theta}^{(t-1)} = (\theta_j^{(t-1)}, \boldsymbol{\theta}_{-j}^{(t-1)})$. The t^{th} iteration of Gibbs sampler comprises M steps, corresponding to the draws from each of the full conditional posterior distributions. Each of these steps can be thought of as a jump in the Metropolis-Hastings sense but with a jumping density that allows only jumps to points in the $\boldsymbol{\theta}$ space that agree with the current point in all components except the one being updated. Let $\theta_{-j}^{(t-1)*}$ denote the subvector of θ that does not contain j^{th} element and has each component set to its most currently updated value. Thus if θ_j, refers to the first element to be updated, $\theta_{-j}^{(t-1)*} = \theta_{-j}^{(t-1)}$, whereas if θ_j is updated at a later step in the updating sequence $\theta_{-j}^{(t-1)*}$ will comprise values updated at the t^{th} iteration for components updated earlier than θ_j in the updating sequence, and values corresponding to the $(t-1)^{th}$ iteration for components of θ updated later in the updating sequence than θ_j.

At the j^{th} step of the t^{th} Gibbs sampler iteration, the full conditional $p(\theta_j^{new}|\boldsymbol{\theta}_{-j}^{(t-1)}, \text{data})$ can be viewed as a jump from $\boldsymbol{\theta}^{(t-1)*}$ to the proposed new point $\boldsymbol{\theta}^{new}$ using a jumping distribution that ensures

$$\boldsymbol{\theta}_{-j}^{new} = \boldsymbol{\theta}_{-j}^{(t-1)*}. \tag{1}$$

This idea can be formalised by defining the jumping distribution at each step of the t^{th} iteration at the Gibbs sampler as follows:
for $j = 1, \dots, M$
draw from

$$J_{j,t}(\boldsymbol{\theta}^{new}|\boldsymbol{\theta}^{(t-1)*}) = \begin{cases} p(\boldsymbol{\theta}_j^{new}|\boldsymbol{\theta}_{-j}^{(t-1)*}, \text{data}) \text{ if } \boldsymbol{\theta}_{-j}^{new} = \boldsymbol{\theta}_{-j}^{(t-1)*} \\ 0 \text{ otherwise} \end{cases} \tag{2}$$

The jumping distribution (2) enforces the condition that at each step j of the t^{th} iteration the only possible jumps are to points which are identical with respect to all elements except those in the j^{th} block.

1.3 Deriving the Metropolis-Hastings acceptance ratio

Suppose θ_j^{new} is a draw from the j^{th} full conditional of a Gibbs Sampler at the t^{th} iteration. The jumping distribution (2) ensures

$$\boldsymbol{\theta}^{new} = (\theta_j^{new}, \boldsymbol{\theta}_{-j}^{(t-1)*}). \tag{3}$$

In addition, since for any j, and any value of $\boldsymbol{\theta}$, $\boldsymbol{\theta} = (\theta_j, \boldsymbol{\theta}_{-j})$

$$p(\boldsymbol{\theta}|\text{data}) = p(\theta_j|\boldsymbol{\theta}_{-j}, \text{data})p(\boldsymbol{\theta}_{-j}|\text{data}). \tag{4}$$

The Metropolis-Hastings ratio at the j^{th} step of the t^{th} iteration is by definition

$$\begin{aligned} r_{MH,j}(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{(t-1)*}) &= \frac{p(\boldsymbol{\theta}^{new}|\text{data})/J_{j,t}(\boldsymbol{\theta}^{new}|\boldsymbol{\theta}^{(t-1)*})}{p(\boldsymbol{\theta}^{(t-1)*}|\text{data})/J_{j,t}(\boldsymbol{\theta}^{(t-1)*}|\boldsymbol{\theta}^{new})} & (5) \\ &= \frac{p(\boldsymbol{\theta}^{new}|\text{data})/p(\theta_j^{new}|\boldsymbol{\theta}_{-j}^{(t-1)*}, \text{data})}{p(\boldsymbol{\theta}^{(t-1)*}|\text{data})/p(\theta_j^{(t-1)*}|\boldsymbol{\theta}_{-j}^{new}, \text{data})} & (6) \end{aligned}$$

But since our jumping density ensures $\boldsymbol{\theta}_{-j}^{new} = \boldsymbol{\theta}_{-j}^{(t-1)*}$ we can replace $\boldsymbol{\theta}_{-j}^{new}$ with $\boldsymbol{\theta}_{-j}^{(t-1)*}$ in the denominator of (6). This gives

$$r_{MH,j}(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{(t-1)*}) = \frac{p(\boldsymbol{\theta}^{new}|\text{data})/p(\theta_j^{new}|\boldsymbol{\theta}_{-j}^{(t-1)*}, \text{data})}{p(\boldsymbol{\theta}^{(t-1)*}|\text{data})/p(\theta_j^{(t-1)*}|\boldsymbol{\theta}_{-j}^{(t-1)*}, \text{data})} \tag{7}$$

Now, since $\boldsymbol{\theta} = (\theta_j, \boldsymbol{\theta}_{-j})$, the joint posterior density, $p(\boldsymbol{\theta}|\text{data})$, can be written as the product of a conditional and marginal density as in (4). Applying the decomposition (4) in the numerator and denominator of (7) gives

$$\begin{aligned} r_{MH,j}(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{(t-1)*}) &= \frac{p(\theta_j^{new}|\boldsymbol{\theta}_{-j}^{new}, \text{data})p(\boldsymbol{\theta}_{-j}^{new}|\text{data})/p(\theta_j^{new}|\boldsymbol{\theta}_{-j}^{(t-1)*}, \text{data})}{p(\theta_j^{(t-1)*}|\boldsymbol{\theta}_{-j}^{(t-1)*}, \text{data})p(\boldsymbol{\theta}_{-j}^{(t-1)*}|\text{data})/p(\theta_j^{(t-1)*}|\boldsymbol{\theta}_{-j}^{(t-1)}, \text{data})} \\ &= \frac{p(\theta_j^{new}|\boldsymbol{\theta}_{-j}^{new}, \text{data})p(\boldsymbol{\theta}_{-j}^{new}|\text{data})/p(\theta_j^{new}|\boldsymbol{\theta}_{-j}^{(t-1)*}, \text{data})}{p(\boldsymbol{\theta}_{-j}^{(t-1)*}|\text{data})} \end{aligned} \tag{8}$$

Since our jumping density requires $\boldsymbol{\theta}_{-j}^{new} = \boldsymbol{\theta}_{-j}^{(t-1)*}$ (8) can be written

$$
\begin{aligned}
r_{MH,j}(\boldsymbol{\theta}^{new}, \boldsymbol{\theta}^{(t-1)*}) &= \frac{p(\theta_j^{new}|\boldsymbol{\theta}_{-j}^{(t-1)*}\text{data})p(\boldsymbol{\theta}_{-j}^{(t-1)*}|\text{data})/p(\theta_j^{new}|\boldsymbol{\theta}_{-j}^{(t-1)*}, \text{data})}{p(\boldsymbol{\theta}_{-j}^{(t-1)*}|\text{data})} \\
&= \frac{p(\boldsymbol{\theta}_{-j}^{(t-1)*}|\text{data})}{p(\boldsymbol{\theta}_{-j}^{(t-1)*}|\text{data})} \qquad (9) \\
&= 1
\end{aligned}
$$

Thus a Gibbs sampler is equivalent to Metropolis-Hastings sampler with jumping distributions defined by the full conditionals as in (2) and acceptance probability equal to 1.