7.3 The multistage Gibbs sampler

There is a natural extension from the two-stage Gibbs sampler to the general multistage Gibbs sampler. Suppose that, for some $p > 1$, the random variable $\mathbf{X} \in \mathcal{X}$ can be written as $\mathbf{X} = (X_1, \ldots, X_p)$, where the X_i's are either unidimensional or multidimensional components. Moreover, suppose that we can simulate from the corresponding conditional densities $f_1, \ldots, f_p$, that is, we can simulate

$$X_i | x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p \sim f_i(x_i | x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$$

for $i = 1, 2, \ldots, p$. The associated *Gibbs sampling* algorithm (or *Gibbs sampler*) is given by the following transition from $X^{(t)}$ to $X^{(t+1)}$:

Algorithm 8 The Multistage Gibbs Sampler

`At iteration` $t = 1, 2, \ldots,$, `given` $\mathbf{x}^{(t)} = (x_1^{(t)}, \ldots, x_p^{(t)})$, `generate`

1. $X_1^{(t+1)} \sim f_1(x_1 | x_2^{(t)}, \ldots, x_p^{(t)})$;
2. $X_2^{(t+1)} \sim f_2(x_2 | x_1^{(t+1)}, x_3^{(t)}, \ldots, x_p^{(t)})$;

$\vdots$

p. $X_p^{(t+1)} \sim f_p(x_p | x_1^{(t+1)}, \ldots, x_{p-1}^{(t+1)})$.

The densities $f_1, \ldots, f_p$ are called the *full conditionals*, and a particular feature of the Gibbs sampler is that these are the only densities used for simulation. Thus, even in a high-dimensional problem, *all of the simulations may be univariate*, which is usually an advantage.

Example 7.4. As an extension of Example 7.1, consider the multivariate normal density

$$(X_1, X_2, \ldots, X_p) \sim \mathcal{N}_p\left(0, (1-\rho)I + \rho J\right), \tag{7.5}$$

where I is the $p \times p$ identity matrix and J is a $p \times p$ matrix of ones. This is a model for *equicorrelation*, as $\operatorname{corr}(X_i, X_j) = \rho$ for every i and j. Using standard formulas for the conditional distributions of a multivariate normal random variable (see, for example, Johnson and Wichern, 1988), it is straightforward but tedious to verify that

$$X_i | x_{(-i)} \sim \mathcal{N}\left(\frac{(p-1)\rho}{1+(p-2)\rho}\bar{x}_{(-i)}, \frac{1+(p-2)\rho-(p-1)\rho^2}{1+(p-2)\rho}\right),$$

where $x_{(-i)} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$ and $\bar{x}_{(-i)}$ is the mean of this vector. The Gibbs sampler that generates from these univariate normals can then be easily derived, although it is useless for this problem (Exercise 7.5). It is, however, a short step to consider the setup where the components of the normal vector are restricted to a subset of $\mathbb{R}^p$. If this subset is a hypercube,

$$\mathfrak{H} = \prod_{i=1}(a_i, b_i)\,,$$

then the corresponding conditionals simply are the normals above restricted to (a_i, b_i) for $i = 1, \ldots, p$ (in which case an exact algorithm such as `sadmvn` can be used). For more complex constraints, a Gibbs sampler is however (almost) required, as exact solutions do not exist. This Gibbs sampler is still based on normal full conditionals, which are now restricted to subsets of the real line and thus easily simulated (Exercise 2.22). ◀

Exercise 7.5 Given the normal target $\mathcal{N}_p(0, (1-\rho)I + \rho J)$:

a. Write a Gibbs sampler using the conditional distributions provided in Example 7.4. Run your `R` code for $p = 5$ and $\rho = .25$, and verify graphically that the marginals are all $\mathcal{N}(0, 1)$.
b. Compare your algorithm using $T = 500$ iterations with `rmnorm` described in Section 2.2.1 in terms of execution time.
c. Propose a constrained subset that is not a hypercube, and derive the corresponding Gibbs sampler. (*Hint*: Consider, for example, a constraint such as $\sum_{i=1}^m x_i^2 \le \sum_{i=m+1}^p x_i^2$ for $m \le p - 1$.)

Models more complex than the one in Example 7.3 can be considered for the normal sampling model, as in the following case.

Example 7.5. A hierarchical specification for the normal model is the *one-way random effects model*. There are different ways to parameterize this model, but a possibility is as follows (see others in Example 7.14 and Exercise 7.24):

$$\begin{aligned}
X_{ij} &\sim \mathcal{N}(\theta_i, \sigma^2), \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i, \\
\theta_i &\sim \mathcal{N}(\mu, \tau^2), \quad i = 1, \ldots, k, \\
\mu &\sim \mathcal{N}(\mu_0, \sigma_\mu^2), \\
\sigma^2 &\sim \mathcal{IG}(a_1, b_1), \quad \tau^2 \sim \mathcal{IG}(a_2, b_2), \quad \sigma_\mu^2 \sim \mathcal{IG}(a_3, b_3).
\end{aligned} \tag{7.6}$$

Now, if we proceed as before and write down the joint distribution from this hierarchy, we can derive the set of full conditionals

$$\begin{aligned}
\theta_i &\sim \mathcal{N}\left(\frac{\sigma^2}{\sigma^2 + n_i\tau^2}\mu + \frac{n_i\tau^2}{\sigma^2 + n_i\tau^2}\bar{X}_i,\ \frac{\sigma^2\tau^2}{\sigma^2 + n_i\tau^2}\right), \quad i = 1, \ldots, k, \\
\mu &\sim \mathcal{N}\left(\frac{\tau^2}{\tau^2 + k\sigma_\mu^2}\mu_0 + \frac{k\sigma_\mu^2}{\tau^2 + k\sigma_\mu^2}\bar{\theta},\ \frac{\sigma_\mu^2\tau^2}{\tau^2 + k\sigma_\mu^2}\right),
\end{aligned}$$

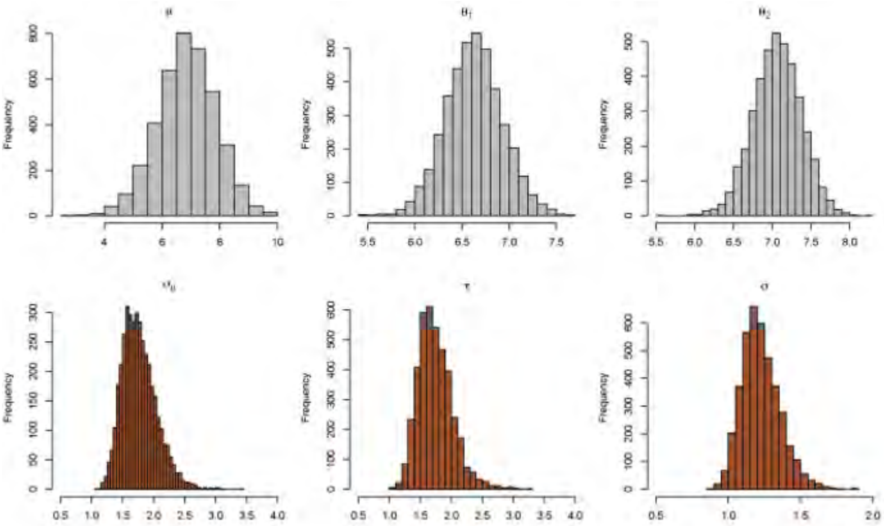

Fig. 7.3. Histograms of marginal posterior distributions from the Gibbs sampler of Example 7.5 based on 5000 iterations. The top row gives histograms for the underlying mean μ and the means, θ_1 and θ_2, for the girls' and boys' energy. The bottom row corresponds to the standard deviations.

$$
\begin{aligned}
(7.7) \qquad \sigma^2 &\sim \mathcal{IG}\left(n/2 + a_1,\ (1/2)\sum_{ij}(X_{ij} - \theta_i)^2 + b_1\right), \\
\tau^2 &\sim \mathcal{IG}\left(k/2 + a_2,\ (1/2)\sum_{i}(\theta_i - \mu)^2 + b_2\right), \\
\sigma_\mu^2 &\sim \mathcal{IG}\left(1/2 + a_3,\ (1/2)(\mu - \mu_0)^2 + b_3\right),
\end{aligned}
$$

where $n = \sum_i n_i$ and $\bar{\theta} = \sum_i n_i \theta_i / n$.

Expanding on the study in Example 7.3, the dataset `Energy` also contains data on the energy intake of boys. Model (7.6) applies (with $k = 2$) to the simultaneous analysis of the energy intakes of girls and boys. The outcome of the Gibbs sampler based on the conditionals in (7.7) is summarized in Figure 7.3. ◀

Exercise 7.6 In the setting of Example 7.5:

a. Derive the full conditional distributions in (7.7).
b. Implement this Gibbs sampler in `R` to reproduce the histograms in Figure 7.3.
c. A variation on the model (7.6) is to give μ a flat prior, which is equivalent to setting $\sigma_\mu^2 = \infty$ in (7.6). Construct the full conditionals for this model and modify the previous `R` code to compare both models on the `Energy` data.