

Practical 2. Markov Chain Monte Carlo

Lecture 2

PDF version

1. Understanding Gibbs sampling

The file [GibbsSampling.xls](#) is a spreadsheet which you can open using MS Excel (or any similar spreadsheet software, such as LibreOffice Calc or OpenOffice Calc) and contains a very simple example of Gibbs Sampling at work. The first worksheet sets up a simple bivariate Normal model:

$$\mathbf{y} = (y_1, y_2) \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\text{with } \boldsymbol{\mu} = (\mu_1, \mu_2) \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}.$$

Given this set up, standard Normal theory says that the *full conditionals* are

- $y_2 \mid y_1 \sim \text{Normal}\left(\mu_2 + \frac{\sigma_2}{\sigma_1}\rho(y_1 - \mu_1), (1 - \rho^2)\sigma_2^2\right)$, and
- $y_1 \mid y_2 \sim \text{Normal}\left(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(y_2 - \mu_2), (1 - \rho^2)\sigma_1^2\right)$.

In order to familiarise yourself with the MCMC process:

1. Inspect cells J4 and K4; the former sets the initial value of y_1 (which is stored in cell B16), while the latter simulates from the conditional distribution of y_2 given the current value for y_1 . (NB: the notation `NORM.INV(rand(), mean, standard_dev)` instructs Excel to simulate a random draw from a Normal distribution with parameters `mean` and `standard_dev`).
2. Move to the second spreadsheet (named 10iters). This shows the (y_1, y_2) plane with the first 10 simulated values; compare the graph with those in the spreadsheets 100iters, 500iters and 1000iters (showing 100, 500 and 1000 iterations of the Gibbs sampler, respectively). Assess convergence on the basis of the graphs. Are 100 iterations enough?
3. Move to the spreadsheet named Tr_y1; this shows the traceplot for y_1 over 1000 simulations. Would you assess that convergence has been satisfactorily reached for this variable? Repeat this for the worksheet Tr_y2, which shows the traceplot for y_2 .
4. Go back to the first spreadsheet and modify the initial values for y_1, y_2 . Go to cell B16 and type the Excel command `=D16` and to cell B17 and type the command `=D17` (these will copy over two random initial values drawn from a Normal distribution with mean 0 and standard deviation 10). Inspect the other spreadsheets; how is convergence affected for (y_1, y_2) ?
5. Go back to the first spreadsheet and modify the value for the correlation coefficient in cell E8, by typing 0.99. Inspect the other spreadsheets; how is convergence affected for (y_1, y_2) ?

⚠ NB: This is not really a problem that requires Gibbs sampling. In fact, that's not even a full Bayesian model (where observed data are used to learn about the model parameters). And more to the point, it's not even a "forward sampling" problem (like the ones we saw in [practical 1](#)), because in this case we're actually fixing the parameters of the model to some fixed values...

Nevertheless, it may be handy to understand and visualise in the spreadsheet the process of repeated sampling from the full conditional distributions that characterises the Gibbs sampling algorithm.

And you should check out the [MCMC tutorial](#) to see the code used to make the "semi-conjugated" example discussed in [class](#).

2. MCMC in BUGS

Consider again the drug example discussed in the lecture. The file [drug-MCMC.odc](#) contains the code to describe the model as well as data and initial values for a simple analysis. From , open the file and, using the commands and actions discussed in the first practical, execute the following analyses.

- Click on Model and then Specification to open the Specification Tool. Check the model (which is contained in the code prefixed by the comment "### Model description", which describes in general the distributional assumptions); load the data (contained in the code prefixed by the comment "### DATA in separate list (better in most cases)"); compile the model; and load the initial values (contained in the code prefixed by the

comment “`### INITIAL VALUES`”). Click on `Inference` and then `Samples` to open the `Sample Monitor Tool`.

Choose the nodes you think should be monitored and then click on `Model` and then `Update` to set up the MCMC run, selecting a suitable number of replications. Finally, produce summary statistics and graphical description of the results.

- You can replicate the above analysis by using the code prefixed by the comment

```
### Data supplied with model description
### (only feasible for very few data points)
```

In this case, you do not need to `load` the data, as they are already given in the model code and thus are automatically loaded when the model is compiled.

- Try and monitor the node `y`. Why do you think does not let you do this?

i If you are using the [binder VM](#), you will also have access to a series of `R` scripts that essentially replicate these analyses using `R` only (without accessing `BUGS`).

It may be helpful to use them, particularly to get even more familiar with the underlying concepts of simulations (which underpin Monte Carlo and Markov Chain Monte Carlo methods). If you want to use them, you can download them from [here](#).

NB: Instructions to install all the relevant packages that are available in the binder VM on your local machine are available [here](#).

