MCMC2: Lab Session 3: SIR-Topics

Theo Kypraios and Phil O'Neill

Background

In Lecture 3 we discussed different topics with regards to the SIR models and in particular we focused on aspects such as what can or cannot be estimated from the data as well as different techniques to improve the mixing of standard MCMC algorithms.

In this lab session we will see in more detail how the inference of the model parameters is affected by the number of removal times which are observed. In addition, we will also implement MCMC algorithms in which we have integrated out the model parameters and investigate if such a strategy improves the mixing of this MCMC algorithm.

Exercises

Exercise 1

Start by downloading from http://www.maths.nott.ac.uk/personal/tk/files/MCMC2-Seattle/Lab-Sessions/3/ two different datasets:

```
dataset_min.txt
dataset_max.txt
```

The first dataset refers to an outbreak where none of the initially susceptible individuals become infected. The second dataset refers to an outbreak where all the initially susceptible individuals became infected some time during the outbreak.

Exercise 2

Fit an SIR model to these datasets assuming that the infectious period follows a Gamma$(2, \gamma)$ as we did in "Lab Session 3". In addition, assume that *a priori*:

$$\beta \sim \text{Gamma}(1, 10^{-3})$$

$$\gamma \sim \text{Gamma}(1, 10^{-3})$$

Draw samples from the posterior distribution of the parameters β and γ. What do you observe? How do your posterior inferences differ from your prior knowledge in both cases? Comment on your results.

What happens if we assume that *a priori* $\beta \sim Exp(1)$?

Exercise 3

We saw in the lecture that by integrating both parameters out and having this distribution as our target (i.e. the distribution we want to draw samples from), this could lead into a more efficient MCMC algorithm.

i. Derive an expression for $\pi(\mathbf{I}|\mathbf{R})$ where

$$\pi(\mathbf{I}|\mathbf{R}) \propto \int_{\beta} \int_{\gamma} \pi(\beta, \gamma, \mathbf{I}|\mathbf{R}) \, \mathrm{d}\gamma \, \mathrm{d}\beta$$

ii. Write a function in `R` which first draws samples from the above target density $\pi(\mathbf{I}|\mathbf{R})$ using MCMC. Then modify this function in order to derive posterior samples from $\pi(\beta|\mathbf{R})$ and $\pi(\gamma|\mathbf{R})$.

iii. Fit an SIR model with Gamma infectious period to the same dataset that we used in Lab Session 2 (Questions 4 and 5). Compare the posterior samples with the samples obtained in Lab Session 2. The answers (eg. the posterior distributions) should be the same. Why?

iv. Does such a strategy improve mixing as compared to the algorithm which does not involve integrating the parameters out? **Hint**: One way to compare the efficiency is to compare the ACF plots of the samples obtained by two different algorithms.