MCMC2: Lab Session 2: Coding and Output

Theo Kypraios and Phil O'Neill

Background

In Lecture 2 we discussed how Bayesian inference can be drawn for the parameters of a stochastic epidemic model using Markov Chain Monte Carlo algorithms. In this lab session we will look at these algorithms in more detail.

1. Start by downloading the file `coding.R` from http://www.maths.nott.ac.uk/personal/tk/files/ MCMC2-Seattle/Lab-Sessions/2/ and save it in your workspace. This file contains three functions which are used to calculate the likelihood of the augmented data (infection and removal times). These functions are needed to build an MCMC algorithm later on (see file `mcmc-Markov.R`):

`count.no.inf`: This function counts the number of infective individuals just before any (arbitrary) time t. Note that an individual, labeled as i, is infective just before time t if $(I_i < t < R_i)$ where I_i and R_i denote their infection and removal time respectively. Therefore if we go through each (ever infected) individual and count how many of them satisfy this condition then we have the desired number.

`compute.total.pressure`: This function computes the integral $\int S_t I_t \, dt$ by making use of the fact that it can be re-written as a double sum as described in Lecture 2.

`compute.log.prod`: This function computes the product which is required for the calculation of the likelihood:

$$\log \left\{ \prod_{j \neq a} I_{i_j^-} \right\}$$

Note that $I_{i_j^-}$ denotes the number of infected individuals just before the j_{th} infection time (i_j^-) and a denotes the label of the initial infective. Note that this function uses the function `count.no.inf`.

2. Download the file `mcmc-Markov.R` from http://www.maths.nott.ac.uk/personal/tk/files/MCMC2-Seattle/ Lab-Sessions/2/ and save it in your workspace.

This file contains one function only, `mcmcSIR.Markov`, which is used to draw posterior samples of the parameters a Markovian SIR model by making use of the functions in the file `coding.R` In brief, a pseudo-code of this MCMC algorithm is given below:

   i. Initialisation;
   ii. Choose one infection and update it using a Metropolis-Hastings step;
   iii. Update β and γ by drawing from their conditional distributions using a Gibbs step;
   iv. Goto to Step (i);

Note that:

- the infection rate (β) and the removal rate (γ) are assumed to be independent and follow *a priori* Gamma distributions with parameters $(\lambda_\beta, \nu_\beta)$ and $(\lambda_\gamma, \nu_\gamma)$, i.e.

$$\pi(\beta) \propto \beta^{\lambda_\beta - 1} \exp\left\{ -\beta \nu_\beta \right\}$$

- the infection times are updated using a M-H algorithm and in particular, using an independence sampler, i.e. by proposing a candidate value, I_i^{can},

$$R_i - I_i^{\mathrm{can}} \sim \mathrm{Exp}(\gamma)$$

Therefore, the q-ratio in the accept/reject probability in the M-H step is given by

$$\frac{(R_i - I_i^{\mathrm{cur}}) \exp\left\{-\gamma(R_i - I_i^{\mathrm{cur}})\right\}}{(R_i - I_i^{\mathrm{can}}) \exp\left\{-\gamma(R_i - I_i^{\mathrm{can}})\right\}}$$

where I_i^{cur} denote the current value of I_i.

Exercises

Start by loading the functions in the files `coding.R` and `mcmc-Markov.R` by using the command `source`, i.e.

```
source("coding.R")
source("mcmc-Markov.R")
```

The function `mcmcSIR.Markov` requires as input a data frame (or matrix) of size $N \times 2$ where N is the size of the population (including the initial infective!). The first column should contain the labels of the individuals and the second column should contain their corresponding removal time. In other words, each row represents an individual. Note that if an individual is known to be susceptible at the end of the epidemic, then their removal and infection time are assumed to be ∞.

Excercise 1

Have a look at the function `mcmc-Markov.R` and **make sure you understand** what is going on.

Excercise 2

Download some simulated data from http://www.maths.nott.ac.uk/personal/tk/files/MCMC2-Seattle/Lab-Sessions/2/ and read them into `R` by using the command:

```
data <-  read.table("data.txt", header=TRUE)
```

Make sure that the data have been read properly and the format is appropriate such that they can be used an an (input) argument to the function `mcmcSIR.Markov`.

Excercise 3

The purpose of this exercise is to fit a non-Markovian model to the observed data where the infectious period is assumed to be Gamma distributed with parameters (α, γ), i.e.

$$R_i - I_i \sim \mathrm{Gamma}(\alpha, \gamma)$$

with $E[R_i - I_i] = \alpha/\gamma$. The parameter α is treated as fixed and known and assume that *a priori* β and γ follow (independent) Gamma distributions as described in Section 1 above.

Before writing any `R` code we need to do write down the density of the posterior distribution we want to draw samples from:

i. Derive the density of the (joint) posterior distribution of the parameters and the infection times given the removal times up to proportionality,

$$\pi(\beta, \gamma, \mathbf{I}|\mathbf{R}) \propto \pi(\mathbf{I}, \mathbf{R}|\beta, \gamma) \times \pi(\beta) \times \pi(\gamma)$$

ii. Derive the densities of the full conditional distributions for the parameters and the unobserved infection times up to proportionality, i.e. $\pi(\beta|\gamma, \mathbf{I}, \mathbf{R})$, $\pi(\gamma|\beta, \mathbf{I}, \mathbf{R})$ and $\pi(\mathbf{I}|\mathbf{R}, \beta, \gamma)$. To do this, simply look at the density of joint posterior distribution, and for example, to derive the density of the full conditional distribution of β then only put together the terms that involve β. What you end up with is the full conditional density you are after (up to proportionality).

Excercise 4

Write a function in R which will draw samples from the posterior distribution of interest, $\pi(\beta, \gamma, \mathbf{I}|\mathbf{R})$, using MCMC and by making use of the function in the file `coding.R`.

Hint 1: Your code should iterate the following steps

i. Choose one infection time and update it using a M-H algorithm;
ii. Update β;
iii. Update γ;
iv. Goto Step (i);

Hint 2: Note that you do not have to write this function from scratch but you can modify the existing function `mcmcSIR.Markov`. However, it might be better to create a new file called `mcmcSIR.gamma` and copy-paste the parts which you can use straight away from `mcmcSIR.Markov`.

Excercise 5

Use the function that you have written in Exercise 4 and use it to fit the above SIR model with Gamma distributed infectious periods to the observed data and draw 10,000 samples from the posterior distribution of the parameters using the function you wrote in part 2).

Assume that $\alpha = 2$. In addition, we assume that we have weak prior information for the parameters and therefore we choose $\lambda_\beta = \lambda_\gamma = 1$ and $\nu_\beta = \nu_\gamma = 10^{-3}$.

- Look at the output by plotting the trace plots of the parameters β, γ and the sum of the infection times $\sum_i I_i$.
- Look at the posterior correlation between β and γ by drawing a scatter plot of the samples against axes β and γ. Furthermore, look at the correlation between the (sum of the) infection times and γ.
- How does the mixing of β and γ compare with the mixing of $R_0 = \beta\alpha/\gamma$? Why?
- Draw a histogram of the posterior distribution of R_0