HA2 - SF2955

Group 2
Magnus Axén maaxen@kth.se 980915-7119
Henrik Fagerlund hfag@kth.se 991012-2655

November 8, 2022

Problem 1

In problem 1 we need convince ourselves that, $(S_t, I_t)_{n\in\mathbb{N}}$ is indeed a Markov chain, i.e. that the new states, S_{t+1}, I_{t+1}, are strictly dependant on the former states S_t, I_t. By use of the equations,

$$\begin{cases} S_{t+1} = S_t - \Delta_t^I \\ I_{t+1} = I_t + \Delta_t^I - \Delta_t^R \\ R_{t+1} = R_t + \Delta_t^R \end{cases} \qquad t \in N$$

it is readily seen, that there is a dependence on S_t, I_t respectively, but the values, Δ_t^I, Δ_t^R need to be investigated. The aforementioned deltas are defined as,

$$\Delta_t^R \sim \text{Bin}\left(I_l, p^{i\to r}\right), \quad \Delta_t^l \sim \text{Neg Bin}(\kappa, \varphi). \tag{1}$$

Here κ and $p^{s\to i}$ are defined as,

$$p_t^{s\to i} = 1 - \exp\left(-\lambda(t)\frac{I_t}{P}\right), \quad \kappa = \left(\frac{1}{\varphi} - 1\right) S_t p_t^{s\to i}.$$

Hence we have strict dependence on the former states and thus have a Markov chain. (The other values are just mere constants). Now let's determine the transition probabilities, i.e.

$$q_\theta\left(s_t, i_\ell; s_{t+1}, i_{\ell+1}\right) = \mathbb{P}_\theta\left(S_{t+1} = s_{\ell+1}, I_{t+1} = i_{t+1} \mid S_l = s_\ell, I_\ell = i_\ell\right)$$

We may rewrite the former to being equivalent to,

$$\mathbb{P}_\theta\left(s_t - \Delta_t^I = s_{t+1}, i_t + \Delta_t^I - \Delta_t^R = i_{t+1} \mid S_l = s_\ell, I_\ell = i_\ell\right).$$

As we have used the conditioning in the probability, we may remove it,

$$\begin{aligned} &= \mathbb{P}_\theta\left(\Delta_t^I = s_t - s_{t+1}, \Delta_t^I - \Delta_t^R = i_{t+1} - i_t\right) \\ &= \mathbb{P}_\theta\left(\Delta_t^I - \Delta_t^R = i_{t+1} - i_t \mid \Delta_t^I = s_t - s_{t+1}\right) \mathbb{P}_\theta\left(\Delta_t^I = s_t - s_{t+1}\right) \\ &= \mathbb{P}_\theta\left(\Delta_t^R = i_t - i_{t+1} + s_t - s_{t+1}\right) \mathbb{P}_\theta\left(\Delta_t^I = s_t - s_{t+1}\right). \end{aligned}$$

Where the distribution of the respective deltas is given by equation 1. We end up with the following transition probabilities,

$$\begin{aligned} &\mathbb{P}_\theta\left(S_{t+1} = s_{\ell+1}, I_{t+1} = i_{t+1} \mid S_l = s_\ell, I_\ell = i_\ell\right) \\ &= \mathbb{P}_\theta\left(\Delta_t^R = i_t - i_{t+1} + s_t - s_{t+1}\right) \mathbb{P}_\theta\left(\Delta_t^I = s_t - s_{t+1}\right). \end{aligned} \tag{2}$$

Where once again Δ_t^R has a binomial distribution and, Δ_t^I has a negative binomial distribution.

Problem 2

To determine the likelihood of $f(\boldsymbol{y} \mid \theta)$, we shall note the following,

$$\begin{aligned} f(\boldsymbol{y} \mid \theta) &= \mathbb{P}\left(S_0 = s_0, I_0 = i_0, S_1 = s_1, I_1 = i_1, \ldots, S_T = s_T, I_T = i_T \mid \theta\right) \\ &= \mathbb{P}_\theta\left(S_0 = s_0, I_0 = i_0, S_1 = s_1, I_1 = i_1, \ldots, S_T = s_T, I_T = i_T\right) \end{aligned}$$

Seeing as $\mathbb{P}_\theta$ is just a conditioning on θ, then we use the chain rule for a joint density and arrive at that the former is equal to,

$$\mathbb{P}_\theta(S_0 = s_0) \cdot \mathbb{P}_\theta(I_0 = i_0 \mid S_0 = s_1), \ldots, \mathbb{P}_\theta(I_T = i_T \mid S_0 = s_0, I_0 = i_0, \ldots\ldots, S_T = s_T)$$

We have previously convinced ourselves that $(S_t, I_t)_{t\in\mathbb{N}}$ is a Markov chain. Therefore, each transition probability does only depend on the previous state and we can then use the expression from problem 1, equation 2, together with the fact that $\mathbb{P}\left(S_0 = s_0, I_0 = i_0 \mid \theta\right) = 1$, to state the following product.

$$f(\boldsymbol{y} \mid \theta) = \prod_{t=1}^{T-1} \mathbb{P}_\theta\left(\Delta_t^R = \Delta_t^i + \Delta_t^s\right) \mathbb{P}_\theta\left(\Delta_t^I = \Delta_t^s\right) \tag{3}$$

Where,

$$\Delta_t^i = i_t - i_{t+1}, \quad \Delta_t^s = s_t - s_{t+1}.$$

Problem 3

We later wish to sample the various variables, $\boldsymbol{\lambda}, \mathbf{t}, p^{i\to r}$. In order to do this, we need to compute the full conditionals $\pi\left(\boldsymbol{\lambda} \mid \mathbf{y}, \mathbf{t}, p^{i\to r}\right)$, $\pi\left(\mathbf{t} \mid \mathbf{y}, p^{i\to r}, \boldsymbol{\lambda}\right)$, and $\pi\left(p^{i\to r} \mid \mathbf{y}, \boldsymbol{\lambda}, \mathbf{t}\right)$ up to some normalising constant.

Before doing so we shall note the following,

$$\pi(\theta \mid \mathbf{y}) \overset{\text{def}}{=} \frac{f(\mathbf{y} \mid \theta)\pi(\theta)}{\int f\left(\mathbf{y} \mid \theta'\right)\pi\left(\theta'\right)d\theta'}. \tag{4}$$

Moreover we have that, $\pi(\theta) = \pi(\boldsymbol{\lambda})\pi(\boldsymbol{t})\pi(p^{i\to r})$. Substituting this into the former expression we arrive at,

$$\pi(\theta \mid y) \propto f(y \mid \theta)\pi(\boldsymbol{\lambda})\pi(\mathbf{t})\pi(p^{i\to r})$$

By conditioning on the two of the variables $\boldsymbol{\lambda}, \mathbf{t}, p^{i\to r}$, we can find the conditional priors to the one not conditioned on. Note moreover when conditioning the π portion of a variable it is just a constant, thus we can use proportionality to simplify those expressions.

- Computing $\pi(p^{i\to r} \mid \mathbf{y}, \boldsymbol{\lambda}, t)$.

  From the Bayesian based equation 4, where the denominator is the normalizing constant, we get,

  $$\pi(\boldsymbol{\lambda}, \mathbf{t}, p^{i\to r} \mid \mathbf{y}) \propto \pi(\boldsymbol{\lambda})\pi(\mathbf{t})\pi(p^{i\to r})f(\mathbf{y} \mid \theta).$$

  Since we have a conditional expression where $\boldsymbol{\lambda}$ and $\mathbf{t}$ is assumed to be known, $\pi(\boldsymbol{\lambda})$ and $\pi(\mathbf{t})$ can be seen as constants and we therefore obtain the following expression,

  $$\pi\left(p^{i\to r} \mid \mathbf{y}, \boldsymbol{\lambda}, \mathbf{t}\right) \propto \pi(p^{i\to r})f(\mathbf{y} \mid \theta).$$

Using definition for the prior $\pi(p^{i\to r})$ together with equation 3, this can be rewritten as the following,

$$= \tfrac{1}{B(a,b)} \left(p^{i\to r}\right)^{a-1} \left(1-p^{i\to r}\right)^{b-1} \prod_{i=1}^{d} \prod_{t=t_i}^{t_{i+1}} \mathbb{P}_\theta \left(\Delta_t^R = \Delta_t^i + \Delta_t^s\right) \mathbb{P}_\theta \left(\Delta_t^I = \Delta_t^s\right)$$

$$\propto (p^{i\to r})^{a-1}(1-p^{i\to r})^{b-1} \prod_{t=0}^{T-1} \mathbb{P}_\theta(\Delta_t^R = \Delta_t^i + \Delta_t^s)$$

Here we used that the negative binomial part, i.e. $\mathbb{P}_\theta(\Delta_t^I = \Delta_t^s)$, is independent of $p^{i\to r}$. Furthermore, we know that Δ_t^R is binomial distributed and we use that fact to rewrite the expression and obtain the final expression,

$$\pi(p^{i\to r} \mid \mathbf{y}, \boldsymbol{\lambda}, \mathbf{t}) \propto (p^{i\to r})^{\sum(\Delta_t^i+\Delta_t^s)+a-1}(1-p^{i\to r})^{\sum(i_t-\Delta_t^i-\Delta_t^s)+b-1}$$

This is identified as the following Beta distribution,

$$p^{i\to r} \mid \mathbf{y}, \boldsymbol{\lambda}, \mathbf{t} \sim \text{Beta}\left(\sum_{t=0}^{T-1} \Delta_t^i + \Delta_t^s - a, \sum_{t=0}^{T-1} i_t - \Delta_t^i - \Delta_t^s + b\right) \tag{5}$$

An example of a similar Beta distribution identification can be seen in Lecture 10 (slide 19).

- Computing $\pi(\boldsymbol{\lambda} \mid \mathbf{y}, \mathbf{t}, p^{i\to r})$.

  By performing similar arguments as before, we obtain,

$$\pi\left(\boldsymbol{\lambda} \mid \mathbf{y}, \mathbf{t}, p^{i\to r}\right) \propto \pi(\boldsymbol{\lambda}) f(y \mid \theta) \propto \prod_{i=1}^{d} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i - 1} e^{-\beta_i \lambda_i} \prod_{t=t_i}^{t_{i+1}} \mathbb{P}_\theta \left(\Delta_t^I = s_t - s_{t+1}\right).$$

  Here d denotes the amount of breakpoints, and t_i, t_{i+1} the corresponding breakpoints. We may ignore the $\mathbb{P}_\theta \left(\Delta_t^R = \Delta_t^i + \Delta_t^s\right)$ portion, as there is no dependence on $\boldsymbol{\lambda}$, and use the proportional argument.

- Computing $\pi\left(\mathbf{t} \mid \mathbf{y}, p^{i\to r}, \boldsymbol{\lambda}\right)$.

$$\pi\left(\mathbf{t} \mid \mathbf{y}, p^{i\to r}, \boldsymbol{\lambda}\right) \propto \pi(\boldsymbol{t}) f(y \mid \theta) \propto \mathbb{1}_{\{0<t_1<t_2<\ldots<t_{d-1}<T\}}(\mathbf{t}) \prod_{i=1}^{d} \prod_{t=t_i}^{t_{i+1}} \mathbb{P}_\theta \left(\Delta_t^I = s_t - s_{t+1}\right).$$

  We ignore the $\mathbb{P}_\theta \left(\Delta_t^R = \Delta_t^i + \Delta_t^s\right)$ terms again as they would be canceled out in the MH algorithm. To more clearly explain the indicator function, we are looking at at the values between the breakpoint above and below.

Problem 4

For the vectors $\boldsymbol{\lambda}$ and $\boldsymbol{t}$ we used a MH algorithm, implemented as seen in the slides of lecture 9. Where the sampling for lambda and t were attained using the random Gaussian walks presented in the lab. As for the $p^{i\to r}$ we used the Gibbs sampler, as presented in lecture 10.

We used logarithmic properties for the MH portion, i.e. for $\pi(\boldsymbol{\lambda} \mid \mathbf{t}, \mathbf{y}, p^{i\to r})$ and $\pi(\mathbf{t} \mid \boldsymbol{\lambda}, \mathbf{y}, p^{i\to r})$, to avoid numerical issues.

Problem 5

The posteriors were investigated to have a very low sensitivity regarding the hyper parameters a and b, and parameters would have to be extremely large to cause any remarkable impact on the result. This being the case, since it only has influence in our beta distribution, in which the values are rather large (considering we are looking at the change and values of infected and susceptible). The algorithmic parameters σ and M causes larger fluctuations to $\boldsymbol{\lambda}$ and $\mathbf{t}$ if increased, but does not impact on the convergence of the values. When increasing σ it was observed a behavior with close to step-like results for $\boldsymbol{\lambda}$, this behavior can be seen in figure 1.

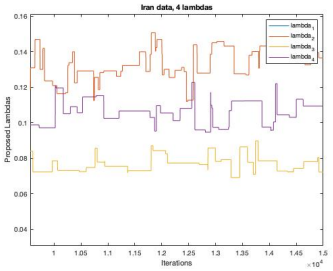


Figure 1: Behavior when using large values for σ.

When investigating $(\beta_i)_{i=1}^d$, it was observed that only β_1 caused any observable impact on the posteriors. It was especially observed on the first λ because of its direct dependence on the prior conditional distribution. When increasing this value, it resulted in a decrease in variance. However, it does also increase the bias and cause the posterior to miss out on relevant relations. For that reason, it can be sub optimal to use a too high/low values for $(\beta_i)_{i=1}^d$.

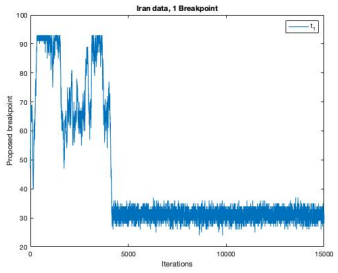
Problem 6

Given the former conditionals, we sample the various variables using MH for $\boldsymbol{\lambda}, \mathbf{t}$ and Gibbs for $p^{i\to r}$ (seeing as we know it's distribution). In the following section we used, $\forall i$, $\beta_i = 3, a = 2, b = 3, M = 2, \sigma = 0.005$. Implementing the former distributions we got the following results.

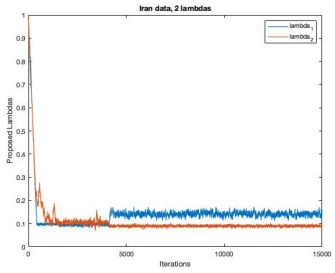
Iran data

We begin by investigating for one breakpoint, by plotting the lambdas and breakpoints estimates. We note for the first plot, figure 2, the breakpoints shift through the plot at two points. This is most likely due to initial indecisiveness of the method. In other words, when we do our random walk, the MH picks up on different breakpoints and wishes to cover all of these rather than just one, causing it to change greatly. Moreover, we can tell as the breakpoint converges, so does the first lambda, jumping up in tune with the breakpoint, causing it to converge.

For two breakpoints, as seen in figure 3, we can tell after some burn in, roughly 1500 or so the breakpoints seem to converge. Moreover we can tell the breakpoints are quite well fitted, in accordance to the infections plot, (in appendix see figure 9). And that the lambdas are the greatest when the infectious population is increasing the most and smallest when it is decreasing.

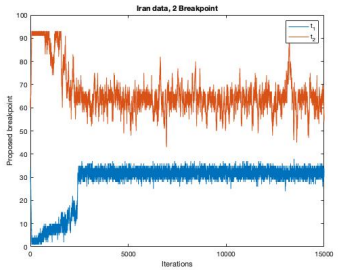


(a) Estimate of breakpoint

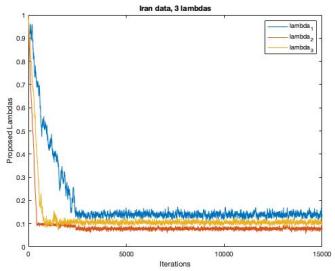


(b) Estimate of Lambdas

Figure 2: Lambda and breakpoint estimate via MH



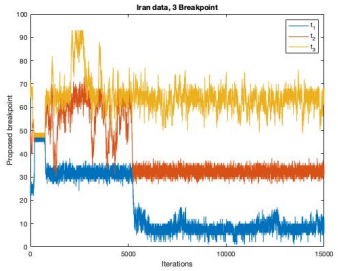
(a) Estimate of breakpoints



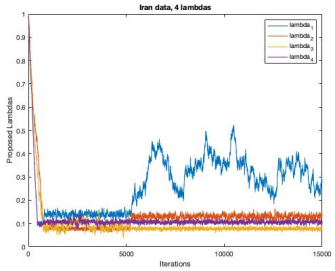
(b) Estimate of Lambdas

Figure 3: Lambda and breakpoint estimate via MH

For the last iteration we got a rather fluctuating lambda for the first breakpoint, we imagine that this has to do with the breakpoint perhaps not being that significant and varying. When having two it was quite clear where those should be (see figure 9) but for three its more unclear.



(a) Estimate of breakpoints



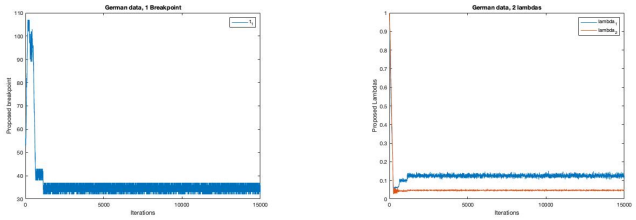
(b) Estimate of Lambdas

Figure 4: Lambda and breakpoint estimate via MH

A mean of $p^{i \to r}$ was computed to be 0.0812, i.e. ≈ 8 %, for $N = 15000$.

German data

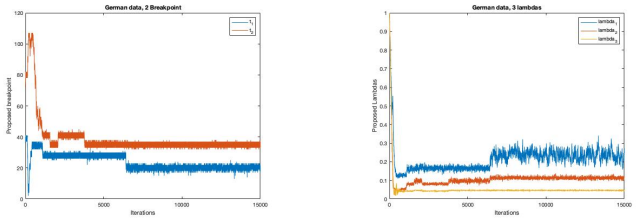
As for the first plot, figure 5, we can see nice convergence for both lambdas, and t. Moreover, it can be seen in figure 10 that our breakpoint is converging towards somewhere around 35 which is roughly where the tipping point is. Similarly here lambda is higher when the infections are increasing and lower when it's decreasing.



(a) Estimate of breakpoint (b) Estimate of Lambdas

Figure 5: Lambda and breakpoint estimate via MH

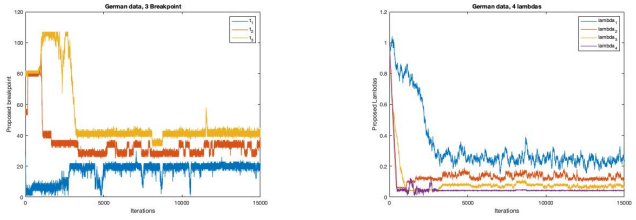
For two breakpoints its dividing up more intervals between the tipping-point and the lambdas are in size according to what we expect. However we notice more variance for the first lambda.



(a) Estimate of breakpoints (b) Estimate of Lambdas

Figure 6: Lambda and breakpoint estimate via MH

In the final plot we see more variance in the breakpoints and again in the first lambda, is varying more.



(a) Estimate of breakpoints (b) Estimate of Lambdas

Figure 7: Lambda and breakpoint estimate via MH

A mean of $p^{i \to r}$ was computed to be 0.0652 , i.e. ≈ 6.5 %, for $N = 15000$.

As to why we see larger variances for some of the lambdas (dependent on the amount of breakpoints) is due to the problems in clear trends. If we inspect the figures in the appendix, we can quite clearly tell that there is two clear breakpoints for the Iran data, and one clear for the German. Thus when these breakpoints are identified the algorithm has trouble identifying the last ones with high certainty, causing larger variations.

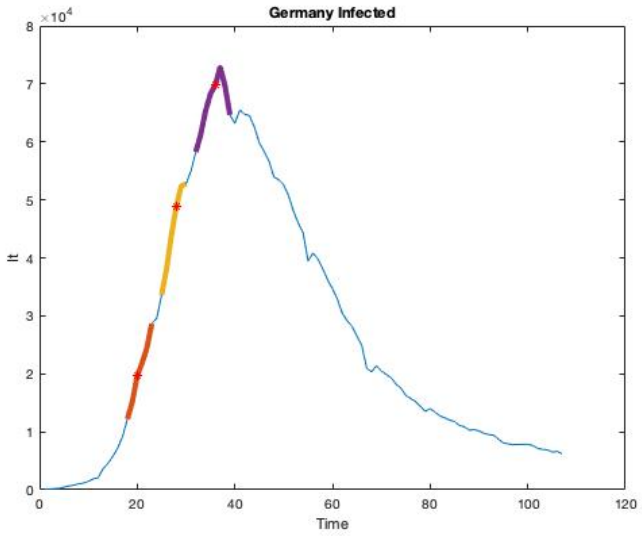


Figure 8: Breakpoints and intervals of the incubation period

In the above figure we have illustrated the different dates when new precautions were taken into consideration, (with the given interval), and plotted the breakpoints from the MCMC, which look to be well placed, given the articles findings.

Problem 7

From equation 5, it can be seen that sampling $p^{i\to r}$ does not require $\boldsymbol{\lambda}$ nor $\mathbf{t}$. Therefore, one would have to sample from a Beta distribution, with parameters solely based on the given data points, and it would not be necessary to use any MCMC algorithm. Even though the data is arrived from an MCMC algorithm this doesn't entail the $p^{i\to r}$ is a Markov chain (and since the data is given).

Appendix

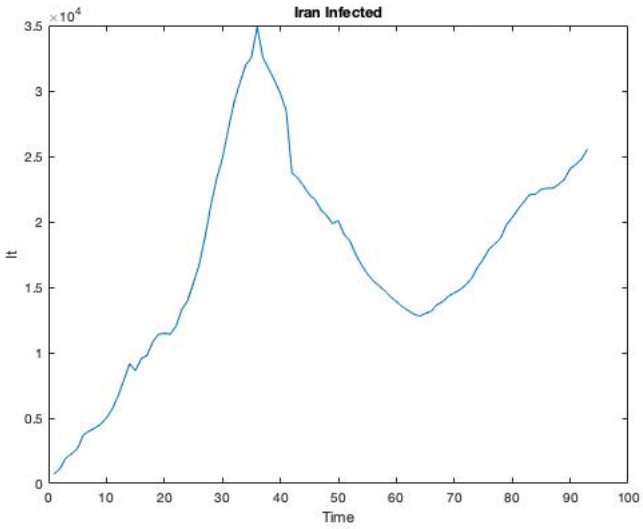


Figure 9: Iran infected

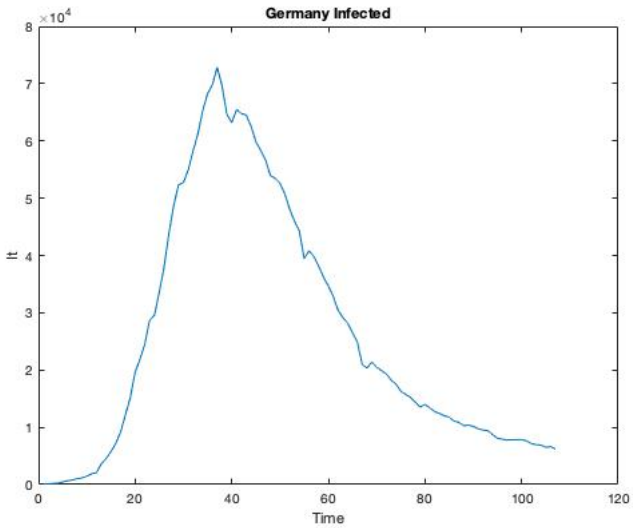


Figure 10: Germany infected