KTH Royal Institute of Technology

SF2955

Computer Intensive Methods in Mathematical Statistics

Home assignment 2

David Ahnlund
000531-4778

Rame Dawli
970720-5333

May 15, 2023

Problem 1

The chain $(S_t, I_t)_{t\in\mathbb{N}}$ is considered as a Markov chain iif the new state of the process (S_{t+1}, I_{t+1}) is only dependent on the previous one (S_t, I_t). Given the equations

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

we certainly see that both S_{t+1} and I_{t+1} depends on the previous states S_t, I_t and two delta factors Δ_t^I and Δ_t^R that we have to study closely. This two delta factors have the following distributions:

$$\Delta_t^I \sim \text{NegBin}(\kappa, \varphi), \qquad\qquad \Delta_t^R \sim \text{Bin}(I_t, p^{i\to r}) \tag{2}$$

where κ and $p^{s\to i}$ are given as following:

$$\kappa = (\frac{1}{\varphi} - 1)S_t p^{s\to i}, \qquad\qquad p^{s\to i} = 1 - \exp(-\lambda(t)\frac{I_t}{P}) \tag{3}$$

Since φ, $\lambda(t)$, and P are only constants that do not affect waiter $(S_t, I_t)_{t\in\mathbb{N}}$ is Markov or not, we see, given (1) and (3), that the new state (S_{t+1}, I_{t+1}) depends only on the system at the previous state (S_t, I_t). Hence, we conclude that $(S_t, I_t)_{t\in\mathbb{N}}$ is a Markov chain. The transition probabilities are given as following:

$$q_\theta(s_t, i_t; s_{t+1}, i_{t+1}) = \mathbb{P}_\theta(S_{t+1} = s_{t+1}, I_{t+1} = i_{t+1} | S_t = s_t, I_t = i_t) \tag{4}$$

$$= \mathbb{P}_\theta(s_t - \Delta_t^I = s_{t+1}, i_t + \Delta_t^I - \Delta_t^R = i_{t+1} | S_t = s_t, I_t = i_t) \tag{5}$$

$$= \mathbb{P}_\theta(s_t - \Delta_t^I = s_{t+1}, i_t + \Delta_t^I - \Delta_t^R = i_{t+1}) \tag{6}$$

Then we can use the definition of conditional probability

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x | Y = y)\mathbb{P}(Y = y) \tag{7}$$

on eq. 6 to get the following:

$$\mathbb{P}_\theta(i_t + \Delta_t^I - \Delta_t^R = i_{t+1} | s_t - \Delta_t^I = s_{t+1})\mathbb{P}_\theta(s_t - \Delta_t^I = s_{t+1}) \tag{8}$$

Then, by substituting the expressions for Δ_t^I and Δ_t^R

$$\Delta_t^I = s_t - s_{t+1} \tag{9}$$

$$\Delta_t^R = i_t - i_{i+1} + \Delta_t^I \tag{10}$$

we finally get the expression

$$q_\theta(s_t, i_t; s_{t+1}, i_{t+1}) = \mathbb{P}_\theta(\Delta_t^R = s_t - s_{t+1} + i_t - i_{t+1})\mathbb{P}_\theta(\Delta_t^I = s_t - s_{t+1}) \tag{11}$$

Problem 2

Want to determine the likelihood:

$$f(\boldsymbol{y}|\theta) = \mathbb{P}(S_0 = s_0, I_0 = i_0, ..., S_T = s_T, I_T = i_T|\theta) \tag{12}$$

As in the assignment description, the assumption is made that $\mathbb{P}(S_0 = s_0, I_0 = i_0|\theta) = 1$

From Problem 1, we already established the following:

$$\mathbb{P}(S_{t+1} = s_{t+1}, I_{t+1} = i_{t+1} | S_t = s_t, I_t = i_t, \theta) = \tag{13}$$

$$= \mathbb{P}(\Delta_t^R = s_t - s_{t+1} + i_t - i_{t+1} | \theta) \mathbb{P}(\Delta_t^I = s_t - s_{t+1} | \theta) \tag{14}$$

Therefore, eq. 12 can be expressed as a multiplicative sequence of eq. 11, due to its Markov property.

$$f(\boldsymbol{y}|\theta) = \prod_{t=1}^{T-1} \mathbb{P}(\Delta_t^R = s_t - s_{t+1} + i_t - i_{t+1} | \theta) \mathbb{P}(\Delta_t^I = s_t - s_{t+1} | \theta) \tag{15}$$

Problem 3

Want to compute the full conditionals

$$\pi(\boldsymbol{t} \mid \boldsymbol{y}, \boldsymbol{\lambda}, p^{i \to r}) \tag{16}$$

$$\pi(\boldsymbol{\lambda} \mid \boldsymbol{y}, \boldsymbol{t}, p^{i \to r}) \tag{17}$$

$$\pi(p^{i \to r} \mid \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{\lambda}) \tag{18}$$

Since we have $\theta = (\boldsymbol{\lambda}, \boldsymbol{t}, p^{i \to r})$, and the parameters are assumed to be independent, we have

$$\pi(\theta) = \pi(\boldsymbol{\lambda}) \pi(\boldsymbol{t}) \pi(p^{i \to r})$$

where the prior distribution for each of the parameters is given as following:

$$\pi(\boldsymbol{t}) \propto \mathbb{1}_{\{0 < t_1 < t_2 < \ldots < t_{d-1} < T\}}(\boldsymbol{t}) \tag{19}$$

$$\pi(\boldsymbol{\lambda}) = \prod_{i=1}^{d} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i - 1} e^{-\beta_i \lambda_i}, \quad (\alpha_i = 2 \; \forall i) \tag{20}$$

$$\pi(p^{i \to r}) = \frac{1}{\mathrm{B}(a, b)} (p^{i \to r})^{a-1} (1 - p^{i \to r})^{b-1}, \;\; (a, b \in \mathbb{N}) \tag{21}$$

– To compute the posterior of $\pi(\boldsymbol{t})$ we have

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

$$= \prod_{t=1}^{T} \mathbb{P}(\Delta_{t-1}^R = s_{t-1} - s_t + i_{t-1} - i_t | \theta) \mathbb{P}(\Delta_{t-1}^I = s_{t-1} - s_t | \theta) \mathbb{1}_{\{0 < t_1 < t_2 < \ldots < t_{d-1} < T\}(\boldsymbol{t})} \tag{22}$$

– For the posterior of $\pi(\boldsymbol{\lambda})$, we have the following:

$$\begin{aligned} \pi(\boldsymbol{\lambda} \mid \boldsymbol{y}, \boldsymbol{t}, p^{i \to r}) &\propto f(\boldsymbol{y} \mid \theta) \pi(\boldsymbol{\lambda}) = \\ &= \prod_{t=1}^{T} \mathbb{P}(\Delta_{t-1}^R = s_{t-1} - s_t + i_{t-1} - i_t | \theta) \mathbb{P}(\Delta_{t-1}^I = s_{t-1} - s_t | \theta) \\ &\qquad \cdot \prod_{i=1}^{d} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i - 1} e^{-\beta_i \lambda_i} \end{aligned} \tag{23}$$

– For the posterior of $\pi(p^{i\to r})$, we have the following:

$$\pi(p^{i\to r} \mid \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{\lambda}) \propto f(\boldsymbol{y} \mid \theta)\pi(p^{i\to r}) =$$
$$= \prod_{t=1}^{T} \mathbb{P}(\Delta_{t-1}^{R} = s_{t-1} - s_t + i_{t-1} - i_t|\theta)\mathbb{P}(\Delta_{t-1}^{I} = s_{t-1} - s_t|\theta) \tag{24}$$
$$\cdot \frac{1}{\mathrm{B}(a,b)}(p^{i\to r})^{a-1}(1-p^{i\to r})^{b-1}$$

Problem 4

We want to implement a Hybrid sampler for the posterior $\pi(\theta \mid \boldsymbol{y})$, which is defined as

$$\pi(\theta \mid \boldsymbol{y}) = \pi(\boldsymbol{t} \mid \boldsymbol{y}, \boldsymbol{\lambda}, p^{i\to r})\pi(p^{i\to r} \mid \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{\lambda})\pi(\boldsymbol{\lambda} \mid \boldsymbol{y}, \boldsymbol{t}, p^{i\to r}) \tag{25}$$

and each of the full conditionals are shown in eq. 22, 23, 24.

As suggested in the assignment description, a Gibbs sampler was used for $p^{i\to r}$ whereas for $\boldsymbol{\lambda}$ and $\boldsymbol{t}$ Metropolis-Hastings was used for the updates. Also, for λ_i, this was updated using Gaussian noise, whereas t_i was updated for each breakpoint using a random-walk proposal:

$$\lambda_i^* = \lambda_i + \sigma\epsilon, \qquad \epsilon \sim N(0,1) \tag{26}$$
$$t_i^* = t_i + \varepsilon, \qquad \varepsilon \in \{\pm 1, ..., \pm(M-1), \pm M\},\ M > 0 \tag{27}$$

When implementing the Hybrid sampler, some simplification for eq 22, 23 and 24 could be made.

For the posterior of $\pi(p^{i\to r})$, eq. 24, we can quickly see that due to Δ_t^I being independent of $p^{i\to r}$, and together with an inspiration from the complex distribution presented in Lecture 10, slide 19 and given as:

$$f(x,y) \propto \frac{n!}{(n-x!)x!} y^{x-\alpha-1}(1-y)^{n-x+\beta-1} \tag{28}$$

we found that the sampling from the conditional distribution is easier which is given by a Beta distribution as following:

$$\pi(p^{i\to r} \mid \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{\lambda}) \propto \prod_{t=1}^{T} \frac{i_t!}{(i_t - x_t!)x_t!}(p^{i\to r})^{x_t+a-1}(1-p^{i\to r})^{i_t-x_t+b-1} \tag{29}$$
$$\sim \mathrm{Beta}(\sum_{t=1}^{T} x_t - a, \sum_{t=1}^{T} i_t - \sum_{t=1}^{T} x_t + b) \tag{30}$$

where x_t is the input $s_{t-1} - s_t + i_{t-1} - i_t$ for the different time steps.

Also for the posterior of $\pi(\boldsymbol{\lambda})$, eq. 23 the expression could be simplified a bit, since the random variable Δ_t^R is independent of $\lambda(t)$.

$$\pi(\boldsymbol{\lambda} \mid \boldsymbol{y}, \boldsymbol{t}, p^{i\to r}) \propto \prod_{t=1}^{T} \mathbb{P}_\theta(\Delta_{t-1}^{I} = s_{t-1} - s_t) \prod_{i=1}^{d} \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i - 1} e^{-\beta_i \lambda_i} \tag{31}$$

Also due to Δ_t^R being independent of $\lambda(t)$ (which depends on $\boldsymbol{t}$), the posterior expression for $\pi(\boldsymbol{t})$ could also be simplified in a similar manner as $\pi(\boldsymbol{\lambda} \mid \boldsymbol{y}, \boldsymbol{t}, p^{i\to r})$, where the probability

$\mathbb{P}(\Delta^R_{t-1} = s_{t-1} - s_t + i_{t-1} - i_t|\theta)$ could be removed in eq. 22. Therefore the final expression used in the Metropolis-Hastings algorithm is as follows:

$$\pi(\boldsymbol{t} \mid \boldsymbol{y}, \boldsymbol{\lambda}, p^{i\to r}) \propto \prod_{t=1}^{T} \mathbb{P}(\Delta^I_{t-1} = s_{t-1} - s_t|\theta)\mathbb{1}_{\{0<t_1<t_2<\ldots<t_{d-1}<T\}}(\boldsymbol{t}) \tag{32}$$

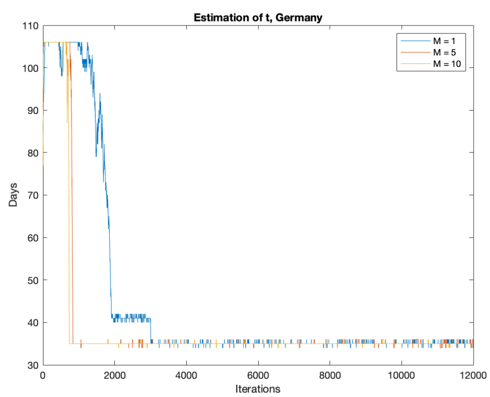
The indicator function, with the condition that the breakpoints ($\boldsymbol{t}$) should be ordered, was determined along with the creating of the new candidate for $\boldsymbol{t}$.

Problem 5

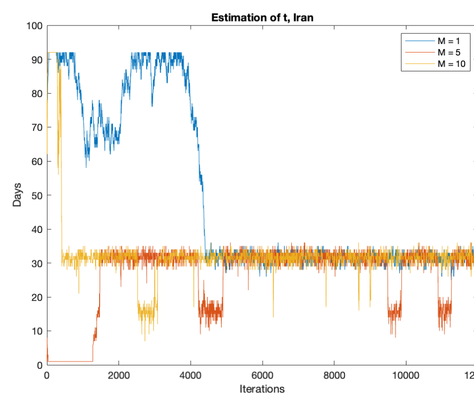
When adjusting the parameters α, β, a, b, σ, and M, it was quickly found that the most sensitive parameter was σ. For increasing σ:s the implementation of the Hybrid sampler became unstable, and for values of $\sigma > 0.05$ the code couldn't execute properly. Therefore, this constant was set to be $\sigma = 0.02$ throughout the entire estimation.

Moreover, it was found that the parameters α, β, a, b did not affect the estimation, nor the mixing, of any noticeable significance.

For the parameter M, i.e. the uniformly drawn stepsize for the breakpoints ($\boldsymbol{t}$) estimation, it was noticeable that for larger values of M the estimated value of $\boldsymbol{t}$ more quickly converged, but was updated less frequently. This phenomenon was however more clear for the Iranian data, as seen in the comparison in Figure 1, where $M \in \{1, 5, 10\}$



(a) Effect on $\boldsymbol{t}$ from M, Germany



(b) Effect on $\boldsymbol{t}$ from M, Iran

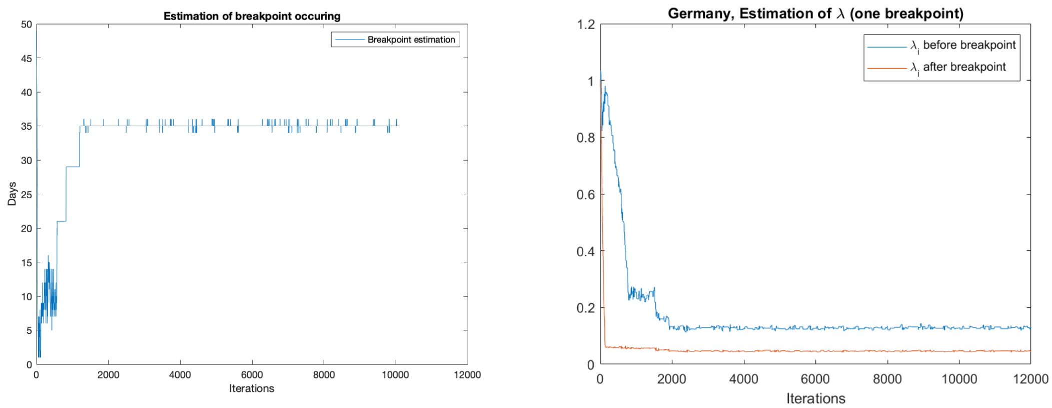
Figure 1: Comparison of M values

Problem 6

For the implementation of the Hybrid sampler, we have assigned these values for the hyper-paramters, $\alpha_i = \beta_i = 2, a = b = 1$ and for the algorithmic parameters $\sigma = 0.02, M = 3$.

Germany

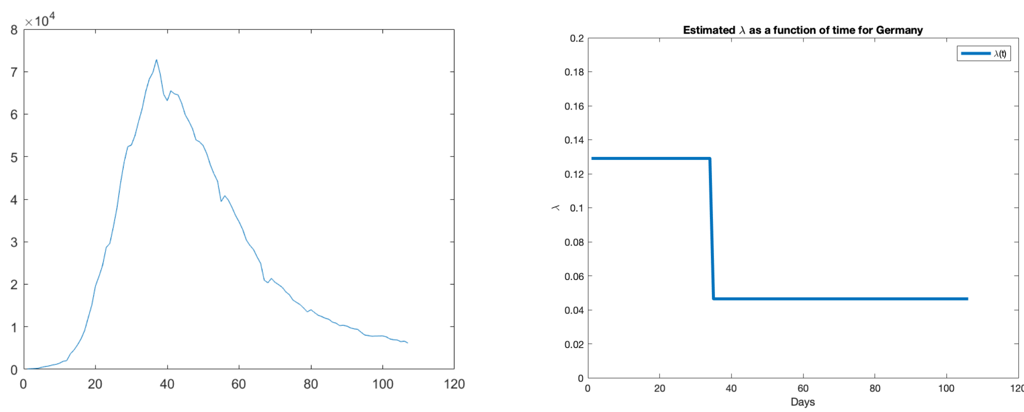
Using the German data, Metropolis-Hasting algorithm gave a steady convergence for both the breakpoint and the lambdas. The breakpoint was estimated to occur at $t = 35.0135$ which we concluded to be a realistic estimation based on the plot of infected people, Figure 3a. We could also observe from Figure 2b that the estimation of $\lambda(t)$ before the breakpoint is higher than after the breakpoint which was also considered as a realistic outcome since the infection is higher when the average number of individual interactions is higher, and vice versa. Meaning that, after the breakpoint the number of infected peopled decreased and due to that $\lambda(t)$ shows lower value after the breakpoint.



(a) Estimation of the breakpoint

(b) Estimation of $\boldsymbol{\lambda}$

Figure 2: Estimation of breakpoint and $\boldsymbol{\lambda}$



(a) Infected Germany

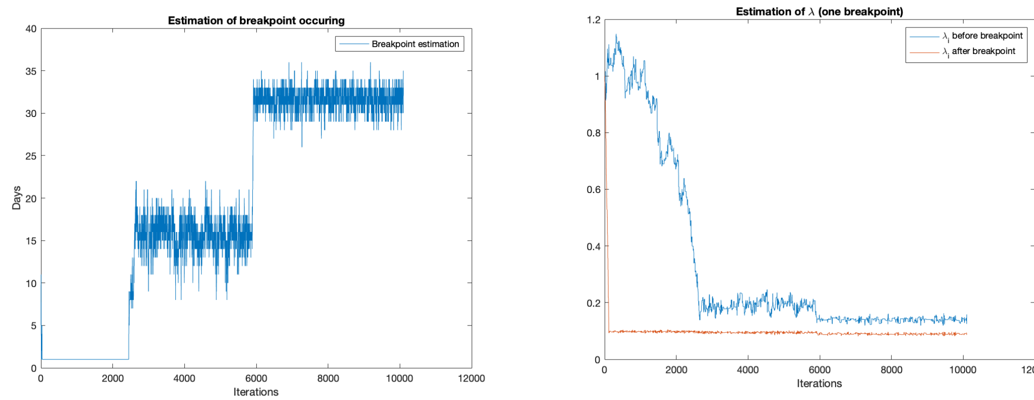
(b) Estimated $\lambda(t)$

Figure 3: Infected Germany and final estimated $\boldsymbol{\lambda}$

Using the Beta distribution of the posterior $\pi(p^{i\to r})$, the mean of $p^{i\to r}$ was estimated to be 0.0695.

Iran

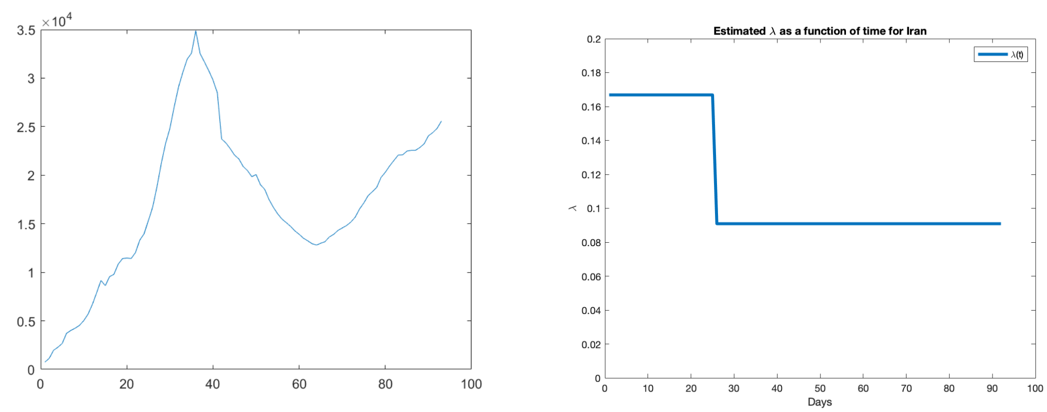
The Iranian data gave more interesting outcomes, which was actually expected by observing the plot for infected people given in Figure 5a. Even if convergence of the breakpoint and the lambdas does not occur as fast as in the German case, Figure 4, we could observe that after enough iterations both of them converges. The estimation of $\boldsymbol{\lambda}$ shows lower values after the breakpoint in this case also as showed in Figure 4b.



(a) Estimation of the breakpoint

(b) Estimation of $\boldsymbol{\lambda}$

Figure 4: Estimation of breakpoint and $\boldsymbol{\lambda}$



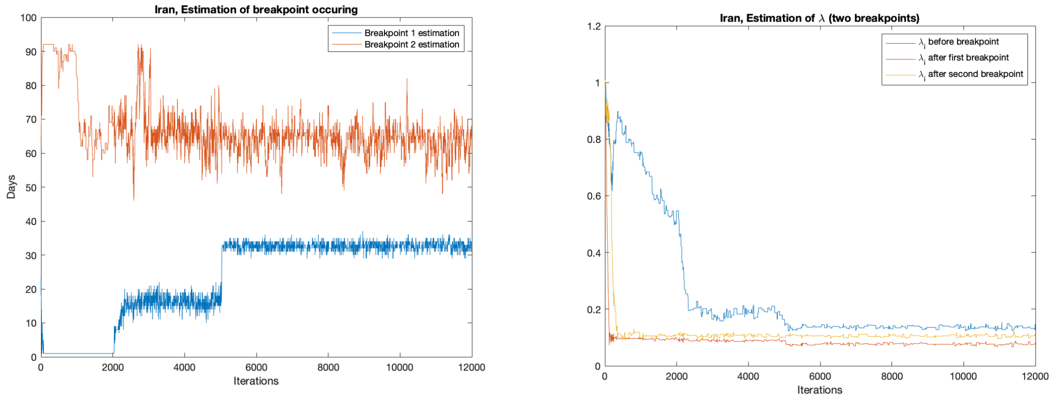
(a) Infected Iran

(b) Estimated $\lambda(t)$

Figure 5: Infected Iran and final estimated $\boldsymbol{\lambda}$

Unlike Germany, the number of infected people in Iran started to increase again after decreasing from all time high. Due to this, the estimation of the breakpoint was harder in this case and the algorithm gave two different estimations for the breakpoint, one around 15 and another one around 33. Looking at Figure 5a we observe that the Iranian data has two different breakpoints where the first one is around 38. However, due to the existence of two breakpoints, the algorithm could not identify the breakpoint with high accuracy when we used only one breakpoint.

Since we could identify two breakpoints in 5a, we implemented the algorithm for 2 breakpoints instead which gave the following results:

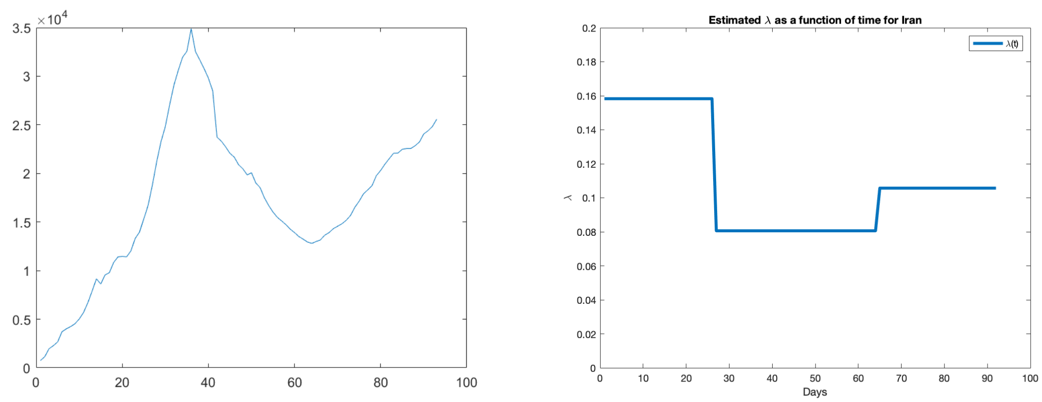


(a) Estimation of the two breakpoints

(b) Estimation of $\boldsymbol{\lambda}$

Figure 6: Estimation of 2 breakpoints and $\boldsymbol{\lambda}$

Figure 6a shows now two estimations for the breakpoints which represents both of the breakpoints in 5a. Compared to the case when we used one breakpoint, we observe now a convergence for the first breakpoint right before $t = 40$ days and the second one around $t = 65$ days.



(a) Infected Iran

(b) Estimated $\lambda(t)$, 2 breakpoints

Figure 7: Infected Iran and final estimated $\boldsymbol{\lambda}$

Using the Beta distribution of the posterior $\pi(p^{i\to r})$, the mean of $p^{i\to r}$ was estimated to be 0.1117.

Problem 7

Looking at the posterior of $p^{i\to r}$ presented in problem 4, we can see that the distribution of $\pi(p^{i\to r} \mid \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{\lambda})$ is independent of both $\boldsymbol{\lambda}$ and $\boldsymbol{t}$. Therefore, if we were interested only in the parameter $p^{i\to r}$ we could sample direct from the Beta distribution given in 30 by using the given data set without the need of the MCMC.