ESPIDAM Stochastic Epidemic Models with Inference – Exercise Session 02

Fanny Bergström

June 25, 2024

Exercise 2.1 (Estimation of R_0)

Assume that a large outbreak occurs in a homogeneously mixing population.

(a)

First assuming there is no preventive measures, estimate the R_0 if we observed that there were 20% infected during the outbreak.

(b)

Suppose now we find out that instead of that each was initially susceptible, there were 70% initially susceptible and the rest 30% were initially immune. We still observed that there were totally 20% infected during the outbreak. Estimate R_0 in this case.

Exercise 2.2 (Estimating parameters in SIR model)

In this exercise, we want to fit the standard SIR model to the data available on our SISMID module website: *covid.csv* (source: `https://covid19.who.int/data`). We will use COVID-19 incidence data from the US during the omicron wave (Dec 2021 – Mar 2022). The US has approximately $N = 331.9$ million population. During this period, there were 28.5 million infections. In the first exercise session (Exercise 1.2), we have solved the SIR differential equation system numerically. Here we consider the least squares corresponding to MLE (Maximum Likelihood Estimates) for Gaussian observations with

$$I(t_i) \sim N(E[I(t_i);\theta], \sigma^2),$$

where $I(t_i)$ the observed number of infected cases at time t_i, the mean $E[I(t_i);\theta]$ determined by the SIR differential equation system and $\sigma^2 = 1$ the fixed variance of observation noise.

(a)

First, define the log-likelihood function using the R command:

```
ll.gauss <- function(theta){
...#determine the solution of SIR ODE
return(sum(dnorm(data, mean =..., sd = 1, log = TRUE)))
}.
```

(b)

Maximize the log-likelihood and compute the corresponding MLE using the R command:

```
optim(..., #initial values for the parameters to be optimized over
fn = ll.gauss, #log-likelihood function
control = list(fnscale = -1) #maximize the function).
```

(c)

Inserting the values of MLE to find the solution of SIR differential equation system. And plot the fitted deterministic curve and the real data of COVID-19 outbreak together. Does it fit well? *Hint*: Try out more starting values, this might be necessary in order to get a quite reasonable fit.

Exercise 2.3 (Estimating parameters in SEIR model)

Our aim here is to fit the SEIR model from Exercise 1.3 with time-varying rate

$$\beta(t) = \begin{cases} \beta_0 \text{ ,if } t \leq t_1 - w, \\ \beta_0 + \frac{\beta_1 - \beta_0}{2w}(t - (t_1 - w)) \text{ ,if } t_1 - w < t \leq t_1 + w, \\ \beta_1 \text{ ,if } t > t_1 + w, \end{cases}$$

to the data of reported COVID-19 cases in Sweden during Nov 2022 – Jan 2023.

We assume that the size of population $N = 11$ million and a latency period of mean 5 days (implying $\rho = 1/5$). Where we set the time $t = 0$ is equal to date 2022-11-14, and let $I(t)$ match the number of reports on calendar day t.
First, optimize the parameters $\theta = (\beta_0, \beta_1, t_1, w, \gamma)'$. Use the simple least-squares approach for fitting and report your estimate $\hat{\theta}$ for θ.

Furthermore, show a plot where you overlay $I(t; \hat{\theta})$ on a time series plot of the number of reported cases per day. Comment your fit. *Hint*: Use the log of the parameters to ensure valid parameter values at all times. Try out more starting values, this might be necessary in order to get a well reasonable fit.