This exercise sheet contains the first of two project works, which needs to be handed in separately by each student signed-up for the course in order to obtain the ECTS credits for it. Your solution shall consist of the following:

- A written report as a PDF file containing solutions in the form of results, textual interpretations and graphs for the four homework exercises. Note: plagiarism or other forms of cheating is a serious act – to underline this your report must as cover page contain the signed confirmation that your work is made in accordance with the Rules for Written Exams at Stockholm University. For further information about possible consequences see also the Guidelines for Disciplinary Matters at Stockholm University.
- Your hand-in is to be uploaded as a bundle consisting of a) scanned copy[1] of your signed `Confirmation.pdf`, b) PDF file `<lastname>.pdf` containing your report c) your commented R/Rmd, Python, Python Notebook, Matlab, Julia, ... file containing the code of your analysis.

  Deadline: **Saturday 01 Aug 2020 at 18:00 o'clock**. All files are to be uploaded before the deadline to the Moodle project 1 task on the course home page. Please note that there is a 10Mb file limit when uploading files. Delayed hand-ins are not accepted.

A total of 25 points can be reached for the answers in the report. Note: A penalty is imposed on reports longer than 20 pages. Your final grade for the project module of the course is determined by your sum of points in the two project works – see the grading criteria of the course for further details.

Lycka till!

Exercise 1 (4 points)

This exercise deals with final size introduced in Lecture 1.

(a) Solve the final size equation $1 - \tau = \exp(-R_0\tau)$ numerically (the largest solution in [0,1]!) as a function of R_0 and create a plot of it for R_0 from 0 to 5.

(b) Do the same thing when a fraction r are immune. However, in the latter case you should plot the overall fraction infected, and not the fraction infected among those who were initially immune.

Exercise 2 (8 points)

In this exercise we consider the Susceptible-Exposed-Infectious-Recovered (SEIR) model in a closed population with parameters $N = 100$, $\beta = 0.004$, $\gamma = 1/7$ and a latency period with mean duration 5 days, i.e. the rate for the $E \to I$ transition is $\rho = 1/5$.

(a) Write up the ordinary differential equation system for the above SEIR model.

(b) Assume one initial infectious at time $t = 0$, i.e. $I(0) = 1$ and use the Euler scheme to find a numerical approximation for $I(t)$, $S(t)$ and $E(t)$ for $t \in [0, 100]$. Show a plot of $S(t)$ and $I(t)$ for this time interval. Hint: You may want to compare your fit with the solution of a more optimized ordinary differential equation solver available in your software package of choice.

(c) Modify the SEIR equations such that $\beta(t)$ becomes a time dependent function, which is β_0 until time $t_1 - w$, is β_1 after time $t_1 + w$ and changes linearly from β_0 to β_1 between $t_1 - w$ and $t_1 + w$. Write the full $\beta(t)$ as part of your report. Hint: Determine what ? should be in the equation below.

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

(d) Let $t_1 = 30$, $w = 5$, $\beta_0 = 0.004$ and $\beta_1 = 0.0012$. Use the Euler scheme or your favorite ODE solver to solve the ODE system numerically and plot $I(t)$ for $t \in [0, 100]$.

[1] Picture taken with a mobile phone is also ok.

Exercise 3 (6 points)

In this exercise you are supposed to fit the SEIR model with time changing transmission rate from exercise 2(c) to the data of reported cases in Stockholm during Feb-Apr 2020.

(a) Fit the deterministic SEIR model with time changing $\beta(t)$ as defined in exercise 2(c) to the data in `Data_2020-04-10Ny.txt` (available from the course webpage). Note: The time series appears to be excluding imported cases. The parameters to optimize for are $\boldsymbol{\theta} = (\beta_0, \beta_1, t_1, w, \gamma)'$. You can use a simple least-squares approach for fitting and pretend that $I(t)$ matches the number of reports on calendar day t. Assumptions: Let $N = 2.37455 \times 10^6$ and let the rate for the $E \to I$ transition be $1/5$. Furthermore,the initial number of infectious is fixed as $I(0) = 1$, where $t = 0$ is equal to 2020-02-17. Report your estimates for $\boldsymbol{\theta}$ and show a plot where you overlay $I(t; \boldsymbol{\theta})$ on a time series plot of the number of reported cases per day. Comment your fit.

Hint: Parametrise your optimization function using the log of the parameters to ensure valid parameter values at all times. Furthermore, hand-tuning the starting values of your optimization might be necessary in order to get a reasonable fit.

(b) Read the FOHM report Estimates of the peak-day and the number of infected individuals during the covid-19 outbreak in the Stockholm region, Sweden February – April 2020[2]. Compile a list of model components, which our above simple SEIR model does not cover, but is included in the FOHM analysis. Note: Difference in the statistical inference methods are not to be considered.

Exercise 4 (7 points)

The file `rT-covid19-deaths-sweden.csv` (available from the course homepage) contains the reporting triangle for all COVID-19 deaths in Sweden as of 2020-06-29. The triangle shows the date of death and the delay until the case appears in the reports of the Folkhälsomyndigheten (FOHM). The data are derived from code made available under an MIT license by Adam Altmejd and are based on the daily FOHM reports. For the simplicity of the analysis in this course we only consider only deaths occuring from 2020-04-02 and onwards. Furthermore, negative increments in the case counts – arising when a death count for particular day is lower in the subsequent FOHM report – are set to zero.

(a) How many deaths are available in the data?

(b) Plot the time series counting the number of deaths per day (i.e. per day of death) as it shows on 2020-06-29 and interpret the time series. Pay particular attention to the time series near time T. Also add a 7-day moving average to the plot, i.e. plot $m(t) = \frac{1}{7}\sum_{i=-3}^{3} y_{t-i}$ on top of the existing plot.

(c) Assume that delays of more than 20 days are irrelevant for the adjustment, i.e. modify the reporting triangle so delays with more than 20 days are put into the 20 bin and reduce the reporting triangle to only contain the delays $0, \ldots, 20$. Furthermore, assume that the delay distribution at time T is stable for the interval of $T, \ldots, T - 20$. We shall restrict attention to this subset of the reporting triangle. Describe a method to estimate the cumulative distribution function of the delay distribution $F(3)$ given that $T =$ 2020-06-29 and state the value you obtain.

Hint: Let D be a random variable representing the reporting delay of a case. The support of D is $0, 1, \ldots, 20$. Consider the probability $g(d) = P(D = d | D \le d)$, i.e. given that we know that $D \le d$, what is the probability that D is exactly equal to D days. How can we estimate $g(d)$ in the reporting triangle? In your calculations you can use without proof that

$$F(d) = P(D \le d) = \prod_{x=d+1}^{20} P(D \le x - 1 | D \le x) = \prod_{x=d+1}^{20} (1 - g(x)).$$

(d) Calculate $F(d)$ for $d = 0, 1, \ldots, 20$ and state the resulting vector.

(e) Explain how you would calculate an estimate for $N(t = \text{2020-06-20}, \infty)$, $\ldots$, $N(t = \text{2020-06-26}, \infty)$ using the above obtained estimate for the delay distribution. State the 7 values.

(f) Compare your results to the graph available from `https://github.com/adamaltmejd/covid/blob/master/docs/archive/deaths_lag_sweden_2020-06-29.png`. Comment the differences.

[2]Code and data available from `https://github.com/FohmAnalys/SEIR-model-Stockholm`