

# Exam: Epidemic modelling

**Candidate nr. 10037**

Department of Physics, Norwegian University of Science and Technology, Trondheim  
Norway

TFY4235 - Computational physics  
(*Last updated on May 7, 2021*)

## Abstract

We study various models for the time evolution of epidemics, including deterministic and stochastic SIR-models, stochastic SEIIaR models and stochastic SEIIaR commuter models. These are applied to a variety of systems, starting small and towards the end applied on a system representing the working population of Norway.

## Contents

<b>I</b>	<b>Introduction</b>	<b>2</b>
0.1	The deterministic SIR model . . . . .	2
0.2	The stochastic SIR model . . . . .	2
0.3	The stochastic SEIIaR model . . . . .	3
<b>II</b>	<b>Code overview</b>	<b>4</b>
0.4	Remarks on performance . . . . .	4
<b>III</b>	<b>Results and Discussion</b>	<b>6</b>
<b>1</b>	<b>Problem 2A: Deterministic SIR model</b>	<b>6</b>
1.1	a) . . . . .	6
1.2	b) . . . . .	6
1.3	c) Flattening the curve. . . . .	7
1.4	d) Vaccinations . . . . .	8
<b>2</b>	<b>Problem 2B: Stochastic SIR model</b>	<b>9</b>
2.1	a) . . . . .	9
2.2	b) . . . . .	10
2.3	c) Probability of an outbreak . . . . .	11
<b>3</b>	<b>Problem 2C: Stochastic SEIIaR model</b>	<b>13</b>
3.1	a) . . . . .	13
3.2	b) Probability of outbreak dependence on $r_s$ . . . . .	15
<b>4</b>	<b>Problem 2D: Stochastic SEIIaR Commuter model</b>	<b>17</b>
4.1	a) Commuter model for a two-town system . . . . .	17
4.2	b) Description of implementation & tests . . . . .	18
<b>5</b>	<b>Problem 2D: Larger stochastic SEIIaR Commuter model</b>	<b>20</b>
5.1	a) 10 city simulation . . . . .	20
5.2	b) 356 city simulation . . . . .	21
5.3	c) Reduced number of commuters in 356 city simulation . . . . .	22

# Introduction

The topic of this project is epidemic modelling. A short introduction to the models under consideration are given in this introduction, mostly for introducing notation and presenting some choices in the implementation. For a more detailed introduction, consult the problem sheet [Nor21].

## 0.1 The deterministic SIR model

The SIR model describes the evolution of **Susceptible**, **Infected** and **Recovered** people in an epidemic. The model reads

$$\frac{dS}{dt} = -\beta \frac{IS}{N} \quad (1a)$$

$$\frac{dI}{dt} = \beta \frac{IS}{N} - I/\tau \quad (1b)$$

$$\frac{dR}{dt} = I/\tau, \quad (1c)$$

where  $\beta$  is the rate at which susceptible people encounter infected people, and become infected, and  $\tau$  is the typical duration of the infection. To solve these equations numerically, we recast it to a system of ODE's by defining the vector  $\mathbf{v} = [S, I, R]^T$ , such that the equations in 1 simplify to

$$\frac{d\mathbf{v}(t)}{dt} = \frac{d}{dt} \begin{bmatrix} S(t) \\ I(t) \\ R(t) \end{bmatrix} = \begin{bmatrix} -\beta IS/N \\ \beta IS/N - I/\tau \\ I/\tau \end{bmatrix} = \begin{bmatrix} -\beta v_1 v_2 / N \\ \beta v_1 v_2 - v_2 / \tau \\ v_2 / \tau \end{bmatrix} = \mathbf{f}(t, \mathbf{v}). \quad (2)$$

## 0.2 The stochastic SIR model

For the stochastic SIR model we have probabilities instead of rates. The probabilities for a susceptible person becoming infected, and an infected person recovering during a time  $\Delta t$  are given by

$$P_{S \rightarrow I} = 1 - \exp(-\Delta t \beta I / N) \quad (3a)$$

$$P_{I \rightarrow R} = 1 - \exp(-\Delta t / \tau), \quad (3b)$$

respectively. The transition to next time step is then governed by

$$S(t + \Delta t) = S(t) - \Delta_{S \rightarrow I} \quad (4a)$$

$$I(t + \Delta t) = I(t) + \Delta_{S \rightarrow I} - \Delta_{I \rightarrow R} \quad (4b)$$

$$R(t + \Delta t) = R(t) + \Delta_{I \rightarrow R}, \quad (4c)$$

$$(4d)$$

where the  $\Delta$ 's are drawn from binomial distributions:

$$\Delta_{S \rightarrow I} = \mathcal{B}(S, P_{S \rightarrow I})$$

$$\Delta_{I \rightarrow R} = \mathcal{B}(I, P_{I \rightarrow R}).$$

To see the connection between this model and the deterministic one, suppose  $\Delta t \beta \ll 1$ , and  $\Delta t / \tau \ll 1$ . In that case, the probabilities in equation (3) are in the first approximation given by

$$P_{S \rightarrow I} \simeq \Delta t \beta I / N \quad (6a)$$

$$P_{I \rightarrow R} \simeq \Delta t / \tau. \quad (6b)$$

In the limit where  $n \rightarrow \infty$ <sup>1</sup>, the binomial distribution with  $n$  samples and probability  $p$  converges to a Poisson distribution with  $\lambda = np$  [Was04], i.e.

$$\Delta_{S \rightarrow I} \rightarrow \text{Pois}(SI\Delta t \beta / N) \quad (7a)$$

$$\Delta_{I \rightarrow R} \rightarrow \text{Pois}(I\Delta t / \tau). \quad (7b)$$

We can find expressions for the  $\Delta$ 's if we assume them to attain the mean value of the Poisson distributions, namely  $\lambda$ :

$$\Delta_{S \rightarrow I} \simeq SI\Delta t \beta / N \quad (8a)$$

$$\Delta_{I \rightarrow R} \simeq I\Delta t / \tau. \quad (8b)$$

Hence, the discrete transition equations (4) are then

$$S(t + \Delta t) - S(t) \simeq -SI\Delta t \beta / N \quad (9a)$$

$$I(t + \Delta t) - I(t) \simeq +SI\Delta t \beta / N - I\Delta t / \tau \quad (9b)$$

$$R(t + \Delta t) - R(t) \simeq +I\Delta t / \tau, \quad (9c)$$

$$(9d)$$

which yields the same equations as the deterministic ones in (1), by dividing all equations by  $\Delta t$  and formally taking the limit  $\Delta t \rightarrow 0$ .

### 0.3 The stochastic SEIIaR model

For the stochastic SEIIaR model we follow essentially the same line of thought as for the stochastic SIR model, except that we include  $E$  : exposed people, that cannot infect others, and  $I_a$  : infected but asymptomatic people. The probabilities for the different transitions happening during an interval  $\Delta t$  are

$$P_{S \rightarrow E} = 1 - \exp \left( -\Delta t \beta \frac{r_s I + r_a I_a}{N} \right) \quad (10a)$$

$$P_{E \rightarrow I} = f_s \times (1 - \exp(-\Delta t / \tau_E)) \quad (10b)$$

$$P_{E \rightarrow I_a} = f_a \times (1 - \exp(-\Delta t / \tau_E)) \quad (10c)$$

$$P_{I \rightarrow R} = 1 - \exp(-\Delta t / \tau_I) \quad (10d)$$

$$P_{I_a \rightarrow R} = 1 - \exp(-\Delta t / \tau_I), \quad (10e)$$

with the different parameters  $r_a, r_s, f_a, f_s, \tau_E, \tau_I$  given in table (1) in [Nor21]. As before,

$$\Delta_{S \rightarrow E} = \mathcal{B}(S, P_{S \rightarrow E})$$

$$\Delta_{I \rightarrow R} = \mathcal{B}(I, P_{I \rightarrow R})$$

$$\Delta_{I_a \rightarrow R} = \mathcal{B}(I_a, P_{I_a \rightarrow R}).$$

---

<sup>1</sup>More formally,  $n \rightarrow \infty$  and  $p \rightarrow 0$  in such a fashion that the product  $np$  is finite.

and

$$\Delta_{E \rightarrow I}, \Delta_{E \rightarrow I_a}, \Delta_{E \rightarrow E} = \mathcal{M}(E, (P_{E \rightarrow I}, P_{E \rightarrow I_a}, (1 - P_{E \rightarrow I} - P_{E \rightarrow I_a}))), \quad (12)$$

where  $\mathcal{M}$  denotes a multinomial distribution, as described in [Nor21]. The changes of the different variables during an interval  $\Delta t$  are then given by

$$S(t + \Delta t) = S(t) - \Delta_{S \rightarrow E} \quad (13a)$$

$$E(t + \Delta t) = E(t) + \Delta_{S \rightarrow E} - \Delta_{E \rightarrow I} - \Delta_{E \rightarrow I_a} \quad (13b)$$

$$I(t + \Delta t) = I(t) + \Delta_{E \rightarrow I} - \Delta_{I \rightarrow R} \quad (13c)$$

$$I_a(t + \Delta t) = I_a(t) + \Delta_{E \rightarrow I_a} - \Delta_{I_a \rightarrow R} \quad (13d)$$

$$R(t + \Delta t) = R(t) + \Delta_{I_a \rightarrow R} + \Delta_{I \rightarrow R}. \quad (13e)$$

$$(13f)$$

To denote a state of the system, I will use  $\mathbf{v} = [S, E, I, I_a, R]^T$ . When we eventually come to the commuter model, I will solve these equations for each group of people in the population structure. In that case, I will refer to  $\mathbf{X}$  as a state of the system, which is the state  $\mathbf{v}$  for all of the groups of people.

PART

II

## Code overview

The source code for this project is structured into separate files for each sub-problem, i.e. `prob_2A.py`, `prob_2B.py` etc. Additionally, there is a file for plotting for each sub-problem. I have also included a file called `utils.py` which contains some useful utility functions used throughout the exercises. For problem 2A I have used the same ODE-solver as the one I made for exercise 2. This can be found in `ode.py`.

### 0.4 Remarks on performance

In many parts I gained a lot of performance by compiling the code with `numba`, as usual. For example, if we solve the stochastic SIR equations with the parameters given in problem 2Ba, and time step  $\Delta t = 0.005$  the performance is as follows:

```
1 # without numba
2 %timeit T,v = stochSIR(v_0,tN,dt,beta,tau)
```

388 ms ± 5.13 ms per loop (mean ± std. dev. of 7 runs, 1 loop each)

To contrast, If we compile both the solver function and the stepper function with `numba` the code runs approximately 50 times as fast:

```
1 # with numba
2 %timeit T,v = stochSIR(v_0,tN,dt,beta,tau)
```

7.46 ms ± 107 µs per loop (mean ± std. dev. of 7 runs, 100 loops each)

My solution for solving the commuter model worked reasonably well in problem 2D, but when I reached problem 2E I found it to be *very* memory consuming, as I store the number

of people in each state for each group of people for each time step, which amounts to an array of size

$$\#\text{time steps} \times \#\text{towns} \times \#\text{towns} \times 5.$$

For the population structure in problem 2Eb this yields an array of approximately 4.7 GiB when using 1000 time steps. To avoid these memory issues I modify the solver to only save the number of people in each state for each group of people for the current and previous time step. This allows for running short time steps without any memory issues, which, importantly, makes it possible to run the 10 simulations in parallel. When naïvely using the solver from problem 2D on the population matrix in problem 2Eb with a step length of 0.1 the magic function `%memit` from IPython produces:

```
peak memory: 10686.46 MiB, increment: 8707.98 MiB.
```

From profiling the code used in problem 2Eb I found that the most time consuming part was the stepping function which draws random numbers for each entry in the matrix. I initially tried to draw these random numbers in a vectorised fashion, but found it exceedingly hard to understand what the functions were doing when providing multidimensional inputs, and I started to get issues with people disappearing from the population. However, I found major improvements just by modifying the line where I step forward using the function `SEIIaR_commuter_step()` to the following:

```
1 if M[1,k] == 0:
2     X[1,k,:] = X_[1,k,:]
3 else:
4     X[1,k,:] = SEIIaR_commuter_step(X_[1,k,:], Pse[k], Pei, Peia, Pir, Piar)
```

This is particularly useful in the case of problem 2Eb and c, as the population matrix is very sparse (see figure 17 for an illustration of the population structure). Testing with and without this change, with a step length of 0.1 yields the following results:

```
1 # Solution without skipping empty entries
2 dt = 0.1
3 %time T, I = SEIIaR_commuter_greedy(M, X_0, tN, dt)
```

```
CPU times: user 1min 17s, sys: 230 ms, total: 1min 17s Wall time: 1min 17s
```

```
1 # Solution _with_ skipping empty entries
2 dt = 0.1
3 %time T, I = SEIIaR_commuter_greedy(M, X_0, tN, dt)
```

```
CPU times: user 22.4 s, sys: 58.5 ms, total: 22.5 s Wall time: 22.4 s
```

A further slight improvement is obtained by only calculating the number of people in each town during daytime and the night once. This brings the runtime slightly lower:

```
CPU times: user 18.2 s, sys: 46.8 ms, total: 18.3 s Wall time: 18.2 s
```

Initially I did this calculation for every iteration, but realised quickly that this was not necessary. I have experienced through this course that the workflow of writing inefficient but nonetheless understandable code to begin with and subsequently trying to optimise it has been way more efficient than trying to find the most optimal solution right away. I find that the example of making this solver faster illustrates this workflow quite nicely.

# Results and Discussion

## 1 Problem 2A: Deterministic SIR model

### 1.1 a)

The ODE-solver I chose to use is the one I made for exercise 2, as I am very familiar with this, and found it working reasonably good for these kind of purposes. For this problem I use the RK4-method. The implementation is an object-oriented version of the solver methods shown in the lectures and should be easily understood by the documentation in `ode.py`.

I solve the deterministic SIR equations in terms of fractions of people. As seen in figure 1 the asymptotic expressions for  $S(t)$  and  $R(t)$  as given in the exam-sheet [Nor21]. The expressions for  $S(\infty)$  and  $R(\infty)$  are solved using the non-linear equation solver `fsove` from `scipy`, with initial guesses 0.5 for both.

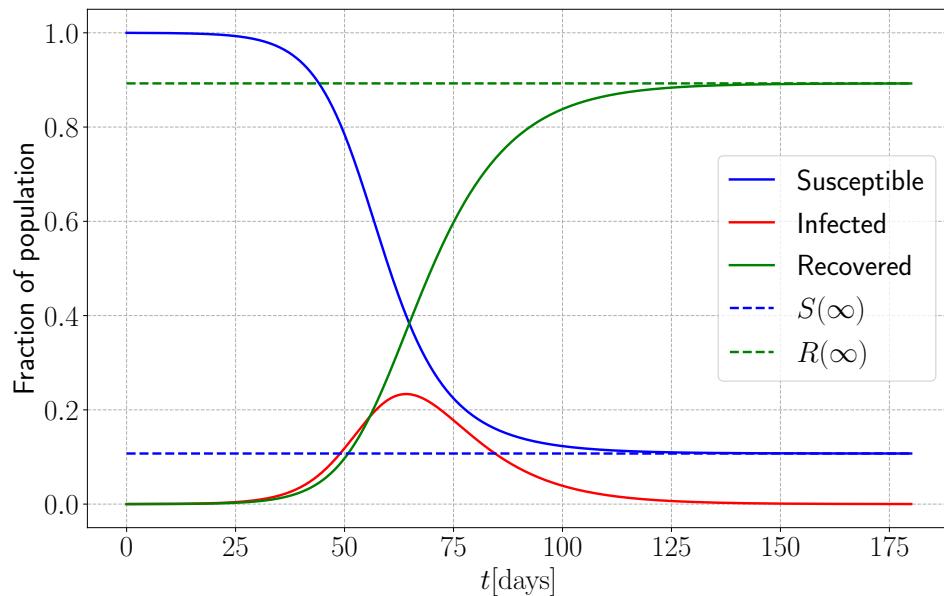


Figure 1: SIR equations with  $\beta = 0.25 \text{ day}^{-1}$ ,  $\tau = 10 \text{ day}$ .

### 1.2 b)

For the early developments of the epidemic, the expression for  $I$  can be simplified to [Nor21]

$$\frac{dI}{dt} = (\beta - 1/\tau)I. \quad (14)$$

The solution of this equation is given by

$$I(t) = I(0) \exp\left([\beta\tau - 1] \frac{t}{\tau}\right) = I(0) \exp\left([\mathcal{R}_0 - 1] \frac{t}{\tau}\right), \quad (15)$$

where we have introduced  $\mathcal{R}_0 = \beta\tau$ . As seen in figure 2, the fraction of infected people matches this expression quite closely during the first 40 days or so. This confirms the exponential growth in the beginning.

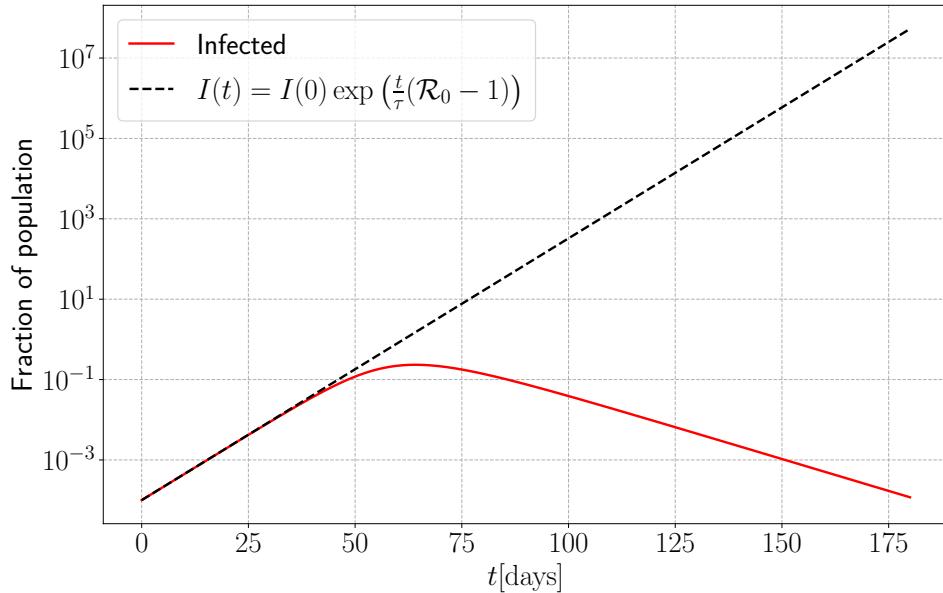


Figure 2: Infected people compared with the analytical approximation at the early stages.

### 1.3 c) Flattening the curve.

As seen from figure 1, the peak of the infected people is rather close to being at 0.2 when  $\beta = 0.25$ . Remark also that the peak occurs in the early stage of the epidemic. We use these observations in our favour when finding the *largest*  $\beta$  that will keep  $I/N^2$  smaller than 0.2 during the pandemic. To do this, I use the following procedure:

```

Choose a starting value for  $\beta = \beta_0$ ;
Choose a tolerance  $\text{tol}$ ;
Run the simulation of the SIR equations with the same parameters as in 2Aa,
except with  $\beta = \beta_0$ ;
Calculate  $\mathcal{I} := \max_{t \in [0, 180]} I(t)$ ;
while  $\text{err} := |\mathcal{I} - 0.2| > \text{tol}$  do
  if  $\mathcal{I} > 0.2$  then
    |  $\beta \leftarrow \beta \cdot 2^{\text{err}}$ 
  else
    |  $\beta \leftarrow \beta \cdot 2^{-\text{err}}$ 
  end
  Run the simulation of the SIR equations with the same parameters as in 2Aa,
  except with the new value for  $\beta$ ;
  Recalculate  $\mathcal{I} = \max_{t \in [0, 180]} I(t)$ ;
end

```

**Algorithm 1:** Finding the largest beta keeping the maximum of  $I$  less than 0.2.

<sup>2</sup>Note that I use fractions of people instead of actual numbers of people in this section, so that is why I refer to  $I$  as if it is  $I/N$  in the procedure below.

This procedure is by no means any advanced piece of algorithm, but it ensures that  $\beta$  is increased when the maximum of  $I$  is less than 0.2, and decreased when it is larger than 0.2. Further, it also ensures that the "nudging" of beta in each step is small when the error is small. To get an estimate for the *largest*  $\beta$  one can have to keep  $I/N$  less than 0.2, one should start off with an initial value  $\beta_0$  for which the peak is less than 0.2, so that one reaches the limit from below. Using  $\beta_0 = 0.2$ , one finds with the maximum value of  $\beta = 0.228020370$ , as given in table 1 together with the deviation  $0.2 - \max_{t \in [0, 180]} I(t)$  this value gives rise to.

## 1.4 d) Vaccinations

To find the minimum fraction of vaccinated people preventing an outbreak,  $R(0)/N$ , I use the following procedure:

```

Choose a starting value for  $R(0) = R_0$ ;
Run the simulation of the SIR equations with the same parameters as in 2Aa,
except with  $R(0) = R_0$  ;
Calculate the slope of the fraction of infected people in a semi-log plot for the
first 15a days of the simulation;
If the initial slope is negative it indicates that the outbreak dies out by itself;
while slope> 0 do
     $R(0) \leftarrow R \cdot 2^{\text{slope}}$  ;
    Run the simulation of the SIR equations with the same parameters as in 2Aa,
    except with the new value for  $R(0)$ .;
    Recalculate the slope in the semi-log axes. ;
end
```

**Algorithm 2:** Finding the minimum fraction of initially vaccinated people for outbreaks to be impossible.

---

<sup>a</sup>As people are typically sick for 10 days ( $\tau = 10$ ) I regard this time scale as long enough to detect whether the outbreak grows exponentially or decays.

The value of  $R(0)/N$  preventing an outbreak found from this procedure is 0.59987499, as shown in table 1 together with the slope in the semi-log plot which this fraction yields for  $I(t)$ . This is obtained from running the procedure with  $R_0 = 0.1$ .

Table 1: The maximum value of  $\beta$  giving a peak less than 0.2 of the infected fraction, and the minimum value of  $R(0)$  (vaccinated) avoiding exponential growth.

Parameter	Value	$0.2 - \max_{t \in [0, 180]} I(t)$	Initial log-slope
$\beta$	0.228020370	$8.319 \cdot 10^{-7}$	—
$R(0)$	0.599874991	—	$-1.74 \cdot 10^{-15}$

## 2 Problem 2B: Stochastic SIR model

### 2.1 a)

As seen in the plot in figure 3, the 10 different realisations of the stochastic SIR model , shown in dashed lines in graded colours, seem to lie close to the deterministic solution.

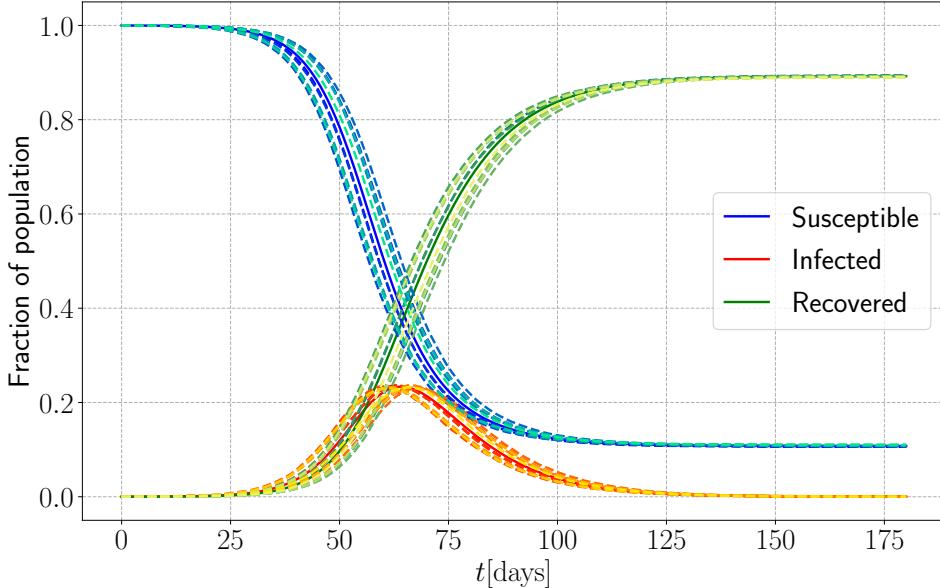


Figure 3: Solution of stochastic SIR equations with  $\beta = 0.25 \text{ day}^{-1}$ ,  $\tau = 10 \text{ day}$ .

To choose a suitable time-step for simulating this model, I run the simulation up to  $t = 10$  days and find the deviation between the fraction of infected people with the semi-analytical result for the early stages given by equation 15. I compute this using the average of 200 runs of the stochastic model. I include the same run with the deterministic model to see how the deviations relate to each other. This is shown in figure 4. Clearly, using this comparison is *not* a good way of determining the time step we want to use for the deterministic model, as the solution approaches the reference values very quickly for the chosen step lengths. The stagnation observed here is probably due to the fact that the reference solution is an approximation itself! However, we see from this plot that if we choose a step length  $\lesssim 0.01$ , the deviations of the stochastic solution start to flatten out, and become smaller than  $10^{-5}$ , which is the highest accuracy we can expect to get when using 100 000 people in the population for the stochastic model (i.e. less than one person difference in the fraction of infected). I therefore always use a time step smaller than 0.01 for simulating the stochastic models. In particular, I use  $\Delta t = 0.005$  in problem B,C,D and Ea. For Eb I use  $\Delta t = 0.01$  because the runtime becomes uncomfortably large when using smaller time steps.

For determining the step length to use for the deterministic model I create a reference solution using a relatively small time step,  $\Delta t = 10^{-4}$ , and then use the deviations between this and each run at the endpoint as a measure of the absolute error. After having done that, it is mostly a matter of preference how small a step length one wants to use. If I can accept an error of  $10^{-8}$  in the fraction of susceptible after 50 days — corresponding to 1/1000 of an individual — I see from the plot in figure 5 that I will have

to use  $\Delta t \lesssim 10^{-1}$ . Since the runtimes of the simulation for  $\Delta t < 10^{-2}$  for 50 days are always less than 1 second, I find that  $\Delta t = 0.005$  is probably more than sufficient, and appropriate considering both runtime and accuracy.

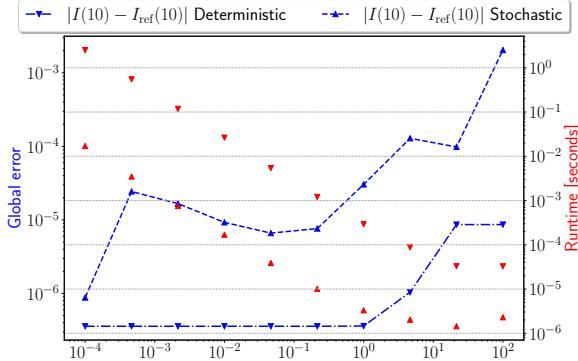


Figure 4: Global error for stochastic model, using the analytical expression for short times,  $I_{\text{ref}}$ , as reference.

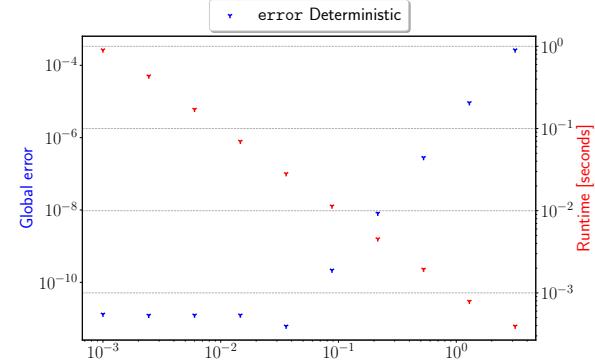


Figure 5: Global error the deterministic model, using a simulation with  $\Delta t = 10^{-4}$  as reference.

## 2.2 b)

As in the previous exercise, we plot the fraction of infected people together with the analytical model for the early stages. This is shown in figure 6. Here we clearly see that all the realisations are approximately linear in the semi-log plot in the first 40 days, or so, of the simulation as we observed in the previous exercise. The mean slope of the 10 realisations is shown in table 2 together with the standard deviation. This is reasonably close to the expected value, which is  $\beta - 1/\tau = 0.15$ .

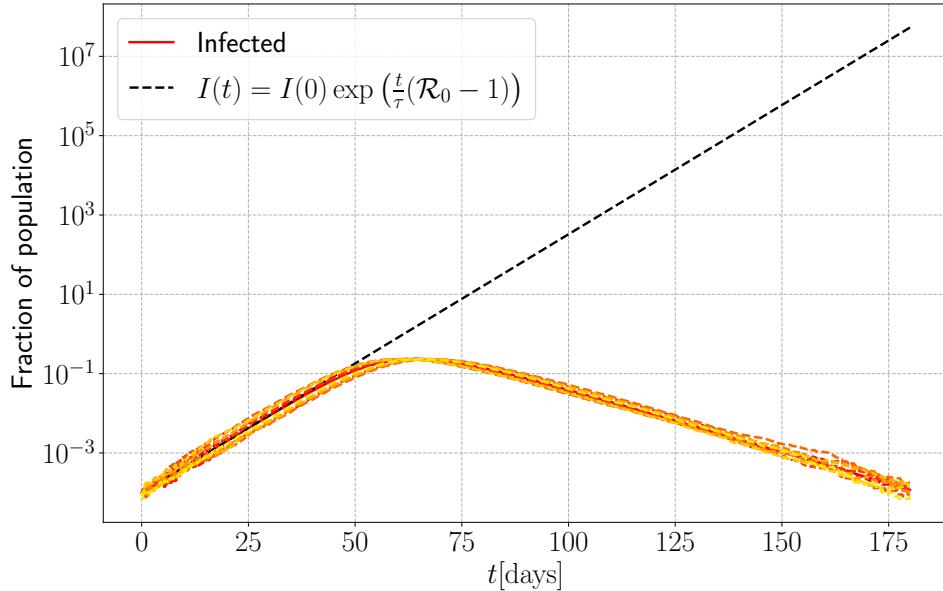


Figure 6: Infected people compared with the analytical approximation at the early stages. Stochastic and deterministic model.

Table 2: Mean slope of the 10 realisation shown in figure 6 and standard deviation.

Initial log slope	Standard deviation
0.1505344881799273	0.013286312850835182

### 2.3 c) Probability of an outbreak

There will always be a certain probability for an outbreak disappearing by itself using the stochastic model. In the present subsection we estimate this probability for an initial number of infected people  $= 1, 2, \dots, 10$ . This is done by the procedure described in algorithm 3. In this procedure, I use the slope of the number of infected people  $I(t)$  during the first 30 days as an indicator of whether the outbreak disappears by itself or not. If the slope is positive, I count it as an outbreak, and if not, I count it as disappearing.

Choose the parameters of the model as those given in the exam sheet [Nor21],  
but with  $T = 30$  days<sup>a</sup> .;

Choose a batch-size  $B$ ;

**for**  $I = 1, 2, \dots, 10$  **do**

Initialise an empty vector of length  $B$ :  $\mathbf{X} = [0, \dots, 0]$  .;

**for**  $n = 1, \dots, B$  **do**

Run the simulation with initial number of infected people  $= I$  .;

Calculate the **slope** in the semi-log axes for  $I(t)$  .;

**if** *slope*  $\leq 0$  **then**

$| X_n = 0$

**else**

$| X_n = 1$

**end**

**end**

Estimate the probability of an outbreak for  $I$  initially infected by

$$p := P(\text{outbreak}|I) = \frac{1}{B} \sum_{n=1}^B X_n.$$

Calculate the standard deviation of the estimate by

$$\sqrt{\text{Var}(\hat{p})} = \sqrt{\frac{p(1-p)}{B}}. \quad (16)$$

**end**

**Algorithm 3:** Calculating the probability of an outbreak as a function of the initial number of infected people,  $I$ .

---

<sup>a</sup>As the typical infection time is 10 days, I assume 30 days to be sufficient for detecting an outbreak in the stochastic model.

In performing this procedure, I use a batch size of 500 in each sweep. What we are estimating here is essentially a Bernoulli-distributed random variable,  $X$ :  $X$  can take the realisations 1 or 0 with probabilities  $p$  and  $1 - p$  respectively, and we assume all the  $X_n$ s to be independent [Was04, p.26]. Then, as the variance of such a random variable is

$p(1 - p)$ , the variance of the estimator for the probability, namely  $\hat{p}^3$ , is

$$\text{Var}(\hat{p}) = \sum_{n=1}^B \text{Var}\left(\frac{X_n}{B}\right) = \frac{1}{B^2} \sum_{n=1}^B \text{Var}(X_n) = \frac{1}{B^2} \sum_{n=1}^B p(1 - p) = \frac{p(1 - p)}{B},$$

from which formula (16) follows.

The probability of an outbreak as a function of the initial number of infected people are shown in figure 7 together with the associated standard deviation. As seen from this plot, when there are more than 6 people initially infected, the probability is approximately 1 that an outbreak will grow. However, for e.g. 1 initially infected person, the probability is approximately  $0.638 \pm 0.021$ . This ultimately shows that the stochastic model has a more realistic feature to it than the deterministic one, in the sense that these scenarios might occur.

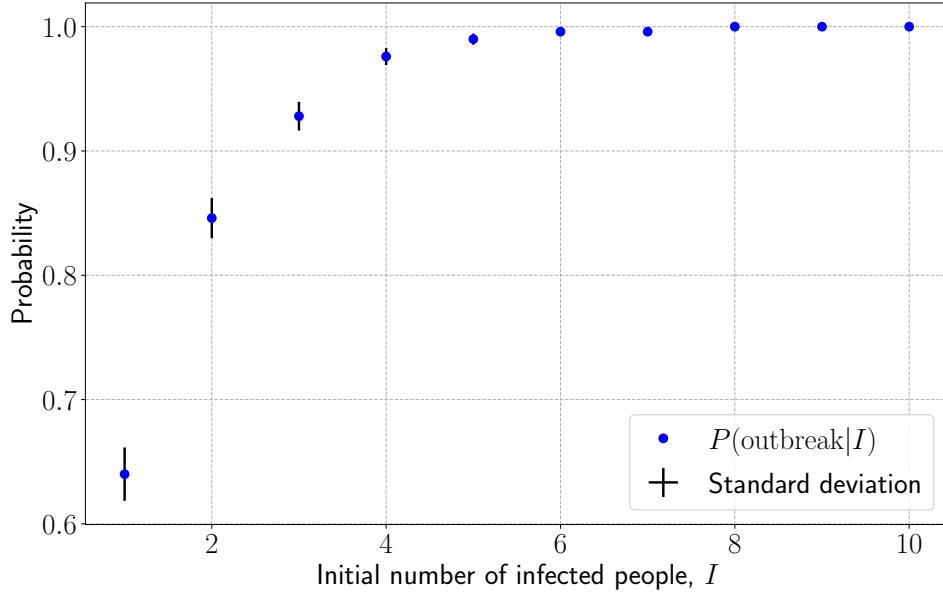


Figure 7: Probability of an outbreak as a function of initial number of infected people.

---

<sup>3</sup>That is, the mean value of  $X_n$ .

### 3 Problem 2C: Stochastic SEIIaR model

#### 3.1 a)

The time-development of all the variables in the SEIIaR model is shown in figure 8, showing 10 realisations of the simulation. The continuous lines shows the solutions of the deterministic model. To compare it easily with the deterministic model, we contract  $E, I, I_a$  into one variable, so that they together represent all the infected people. The solutions of the deterministic model are shown together with the contracted versions of these simulations in figure 9.

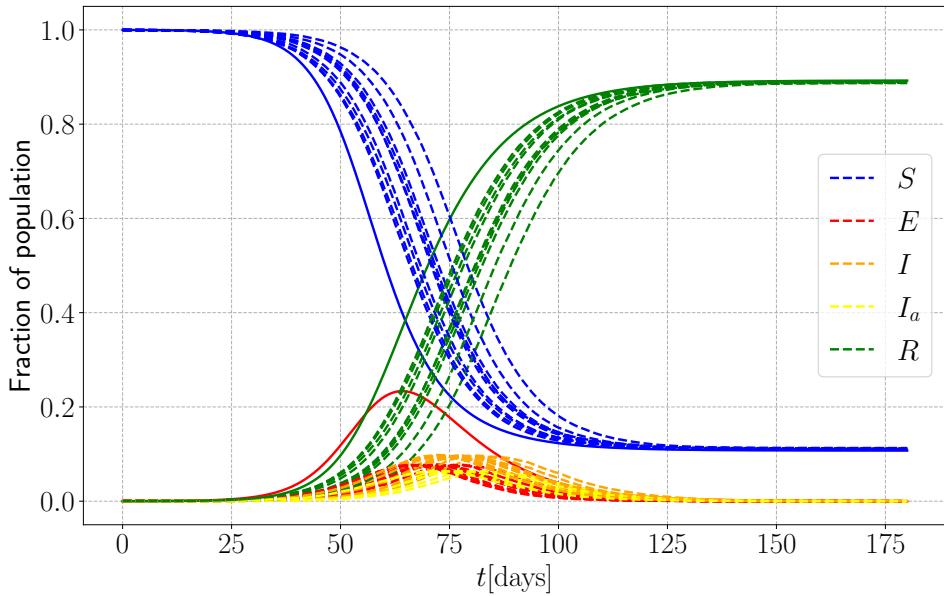


Figure 8: Solution of the stochastic SEIIaR-equations.

One conspicuous feature to notice here, is that the process overall seems to be *slower*, in the sense that it takes longer time for people to get infected and eventually recover, but the asymptotic behaviours seem to align quite closely. This is probably due to the fact that this model includes a period in which people are infected but not yet able to infect anybody else: an incubation period of typical length  $\tau_E = 3$  days. This will naturally delay the process, but it should not affect the asymptotic behaviour. The peak of the infected fraction is also seen to be slightly lower. This is also a result of this delay period making the infections more spread out in time. Ultimately, this shows that the SEIIaR model is more realistic in the sense that people do not get sick right away, and that the behaviour agrees with our expectations. Another thing to mention is that the fraction of infected asymptomatic people  $I_a$  are always less than the fraction of infected symptomatic people  $I$ . This is as expected since  $f_a < f_s$  and  $r_a < r_s$ .

To test that the implementation is correct — by comparing with the deterministic and stochastic SIR model — we adjust the parameters of the SEIIaR to  $\beta = 0.25$ ,  $r_s = 1$ ,  $r_a = 1$ ,  $\tau_E = 0$ ,  $\tau_I = 10$  and start with 10 out of 100 000 people initially infected. The results of doing 1000 such simulations and performing the average of the stochastic models is shown in figure 10. This shows that the two stochastic models are very similar, as expected. Furthermore we see that there seems to be a slight delay of the (mean of

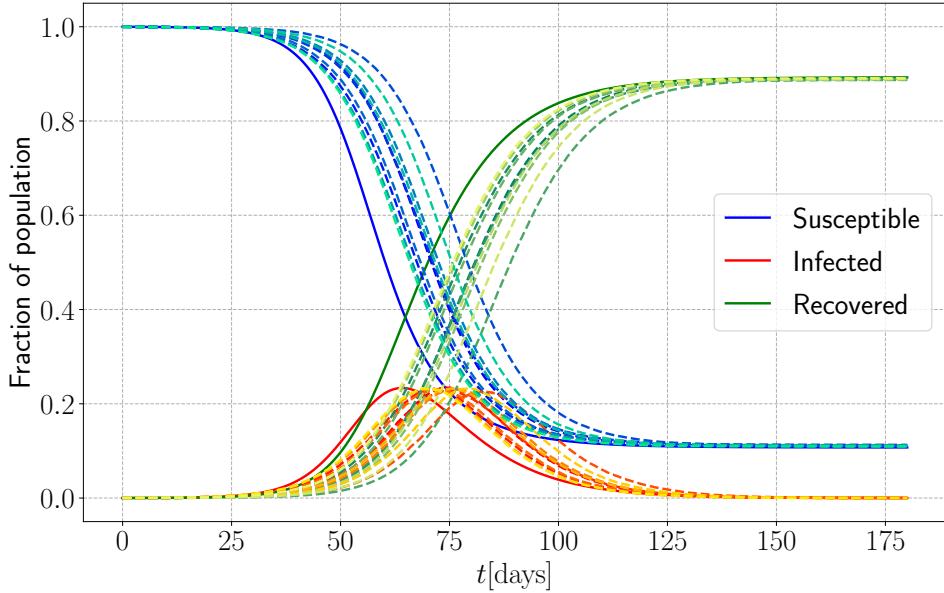


Figure 9: Comparison of the solution of the Stochastic SEIIaR-equations with the deterministic SIR-model. The number of infected people  $I$  in the stochastic model is  $E + I + I_a$ .

the) stochastic models compared to the deterministic one. This is however as expected, since only in the limit of  $\Delta t \rightarrow 0$  and (crucially)  $N \rightarrow \infty$  does the stochastic SIR model coincide with the deterministic one, as discussed in the introduction in section 0.2. Since we use populations of size  $N = 10^5$  we must naturally expect such a deviation, despite the fact that we are using a *small* time step and take the average of many simulations. This also emphasises the fact that using the deterministic model as a reference solution for determining the step length for the stochastic model(s) is probably not a good idea, since we don't even expect them to be equal when  $\Delta t \rightarrow 0$ , for finite  $N$ .

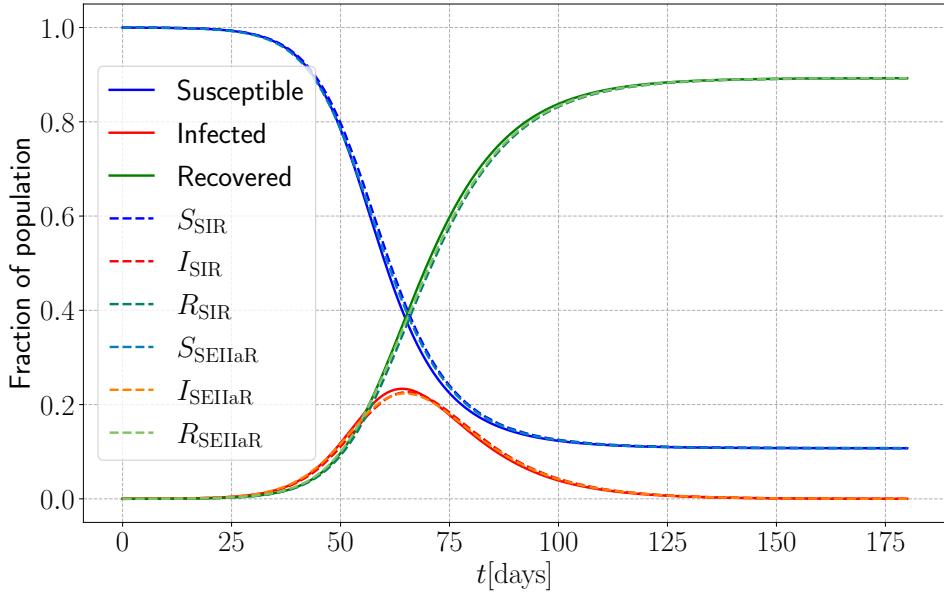


Figure 10: Solution of the stochastic SEIIaR-equations compared with the stochastic and deterministic SIR-equations for the case of identical parameters.

### 3.2 b) Probability of outbreak dependence on $r_s$

The variable  $r_s$  describes how infectious a person is when he is in the infected state. Reducing this constant below 1 can therefore correspond to increasing the degree of self-isolation when people are symptomatic. We investigate the probability of an outbreak as a function of this self-isolation-rate  $r_s$  by the following procedure (which is essentially the same as that shown in 2Bb expect we are now finding the probability as a function of  $r_s$ ):

Choose the parameters of the model as those given in the exam sheet [Nor21],

but with  $T = 30$  days<sup>a</sup> . ;

Choose a batch-size  $B$ .;

Choose  $n$  number of values of  $r_s$  to try;

Select  $n$  values of  $r_s$ , equally spaced between 0.001 and 1 :

$$\mathbf{R} = [R_1, \dots, R_n]$$

**for**  $i = 1, 2, \dots, n$  **do**

Initialise an empty vector of length  $B$ :  $\mathbf{X} = [0, \dots, 0]$ .;

**for**  $n = 1, \dots, B$  **do**

Run the simulation with  $r_s = R_{i,:}$ ;

Calculate the slope in the semi-log axes for  $I(t)$ .;

**if**  $slope \leq 0$  **then**

|  $X_n = 0$

**else**

|  $X_n = 1$

**end**

**end**

Estimate the probability of an outbreak when  $r_s = R_i$  by

$$p := P(\text{outbreak}|r_s) = \frac{1}{B} \sum_{n=1}^B X_n.$$

Calculate the standard deviation of the estimate by

$$\sqrt{\text{Var}(\hat{p})} = \sqrt{\frac{p(1-p)}{B}}.$$

**end**

**Algorithm 4:** Calculating the probability of an outbreak as a function of  $r_s$ .

---

<sup>a</sup>As the typical infection time is still 10 days, I assume 30 days to be more than sufficient for detecting an outbreak in the stochastic model.

The results of this calculation using a batch size of  $B = 500$  and 100 values of  $r_s$  are shown in figure 11. The behaviour is as expected: when people are hardly infectious when symptomatic the probability of an outbreak is close to 0. This is also probably also a result of the fact that  $r_a = 0.1$ , i.e. it is not particularly likely to infect anyone when you are asymptomatic. If  $r_a$  and  $f_a$  was higher, one would expect the probability distribution to stagnate at some finite value when  $r_s \rightarrow 0$ , or at least go to 0 slower. This

is demonstrated in the same figure, where we set  $r_a = 1$  and perform the same test as above. This again is an intuitive confirmation that the model behaves as expected, and therefore is correctly implemented.

We see from this plot that an outbreak is almost certain to *not* grow exponentially when  $r_s \lesssim 0.2$ , and for  $r_s \simeq 0.38$  there is a probability of  $0.46 \pm 0.022$  for an outbreak growing.

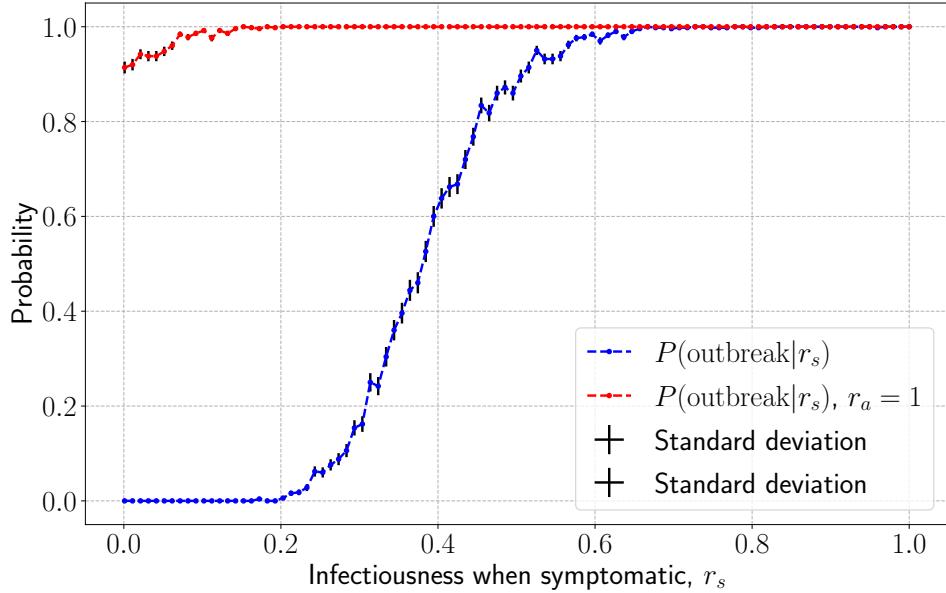


Figure 11: Probability of an outbreak as a function of  $r_s$ .

## 4 Problem 2D: Stochastic SEIIaR Commuter model

### 4.1 a) Commuter model for a two-town system

I set up a population structure as described by the matrix

$$\mathbf{M} = \begin{bmatrix} 9000 & 1000 \\ 200 & 99800 \end{bmatrix},$$

and simulate the time evolution of each of the five states for 180 days, as described in [Nor21], with 25 initially exposed people working and living in town 1. The results of this simulation is shown in figure 12. Notice that we here use the number of people as scale on the vertical axis, so that we easier can distinguish the two towns from each other. We observe that the evolution of the epidemic in town 2 is delayed compared to that of town 1. Notice that people start to get exposed in town 2 roughly around the time at which the number of infected in town 1 reaches its peak. This is as expected, as only 25 people in town 1 are exposed, and they can only infect people in town 2 *via* someone that works in town 2 during the night, or someone that works in town 1 while living in town 2 during daytime. Also here, note that at each time there are fewer asymptomatic infected than symptomatic infected. This is as expected since  $f_a < f_s$ , and  $r_a < r_s$ .

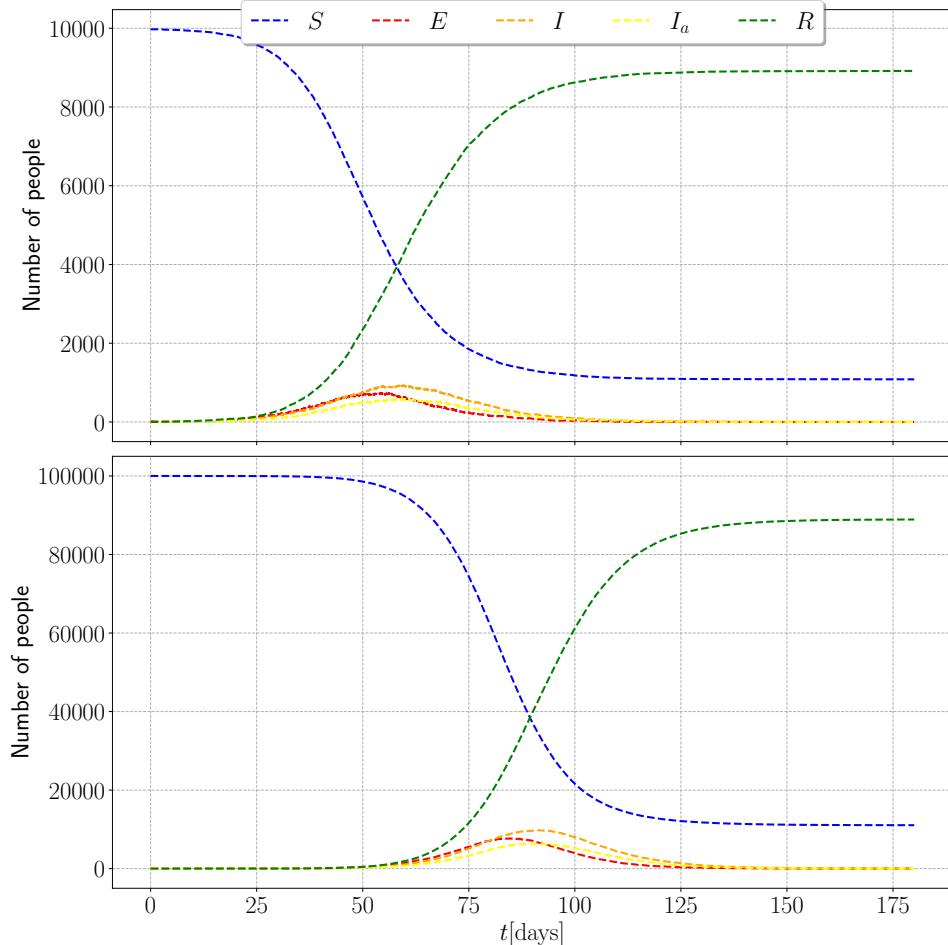


Figure 12: Solutions of Stochastic SEIIaR commuter model for the 2-city case.

## 4.2 b) Description of implementation & tests

To implement the Stochastic SEIIaR Commuter model, I choose the following procedure:

- Choose a population structure  $\mathbf{M} \in \mathbb{N}^m \times \mathbb{N}^m$ , i.e. an  $m \times m$  matrix;
- Set an end time  $t_N$  and a time-step  $\Delta t$ ;
- Set an initial state  $\mathbf{X}_0 \in \mathbb{N}^m \times \mathbb{N}^m \times \mathbb{N}^5$ , where entry  $X_{0,ij}$  is the vector  $\mathbf{v}$  of the variables  $S, E, I, I_a, R$  for group  $(i, j)$  in the matrix  $\mathbf{M}$ ;
- Create an empty array for holding the state at each point in time,  $\mathbf{X}$ , and set its first element to  $\mathbf{X}_0$ ;
- Calculate the probabilities:

$$\begin{aligned} P_{E \rightarrow I} &= f_s \times (1 - \exp(-\Delta t/\tau_E)) \\ P_{E \rightarrow I_a} &= f_a \times (1 - \exp(-\Delta t/\tau_E)) \\ P_{I \rightarrow R} &= P_{I_a \rightarrow R} = 1 - \exp(-\Delta t/\tau_I) \end{aligned}$$

Calculate the number of days  $D = \text{int}(t_N)$ , and the number of steps per half day  $S = \text{int}(1/(2\Delta t))$ ;

**for**  $d = 1, \dots, D$  **do**

$i \leftarrow 2 \times (d - 1) \times S$ ;

**for**  $j = 1, \dots, S$  **do**

Calculate the number of people in each town,  $\beta$ ,

$$N_\beta = \sum_{\alpha=1}^m M_{\alpha\beta}$$

Find the number of infected people of each kind at the previous time step for each town  $\beta$ ,

$$I_\beta, I_{a,\beta} = \sum_{\alpha=1}^m X_{(i+j)\alpha\beta 3}, \sum_{\alpha=1}^m X_{(i+j)\alpha\beta 4}$$

Calculate the probability for transitioning between  $S$  and  $E$  for each town,  $\beta$ , <sup>a</sup>

$$P_{S \rightarrow E, \beta} = 1 - \exp\left(-\Delta t \beta_0 \frac{r_s I_\beta + r_a I_{a,\beta}}{N}\right)$$

**end**

Do a normal SEIIaR step for each group of people  $\alpha, \beta = 1, \dots, m$ , exactly as described in the introduction to update to the present state  $X_{(i+j+1)\alpha\beta 3}$ ;

$i \leftarrow i + S$ ;

**for**  $j = 1, \dots, S$  **do**

Repeat the procedure in the loop above, but with  $\alpha \leftrightarrow \beta$ , i.e. perform the sums over the *third* not *second* axis of  $\mathbf{X}$ , and the *second* of  $\mathbf{M}$ ;

**end**

**end**

**Algorithm 5:** Description of implementation of the SEIIaR commuter model.

---

<sup>a</sup>Note that here I have used  $\beta_0$  to denote the beta value describing the value of the model, to not confuse it with the summation variables.

For more details on exactly how this is implemented in python I refer you the code in `prob_2D.py`.

To check that the system behaves as expected, we simulate the same scenario, only with another matrix which now represents *no* flow of workers to the other areas during daytime:

$$\widetilde{\mathbf{M}} = \begin{bmatrix} 10000 & 0 \\ 0 & 100000 \end{bmatrix}. \quad (17)$$

The time-evolution of the different variables are shown in figure 13, in which it is apparent that the exposed people in the small city never infect those in the large city, as we expect. The expected behaviour is also observed when the exposed start out in area 2, but a plot of this is not included, for brevity.

To do a second test we let all the 10000 of the people living in town 1 work in town 2. That is, we use the population matrix in equation (18)

$$\overline{\mathbf{M}} = \begin{bmatrix} 0 & 10000 \\ 0 & 100000 \end{bmatrix}. \quad (18)$$

We simulate the same situation, with 25 of the 10000 people initially exposed. This should make the evolution of the variables approximately in sync. This is indeed what we observe, in figure 14, as the number of infected people in each town are more in sync than in figure 12.

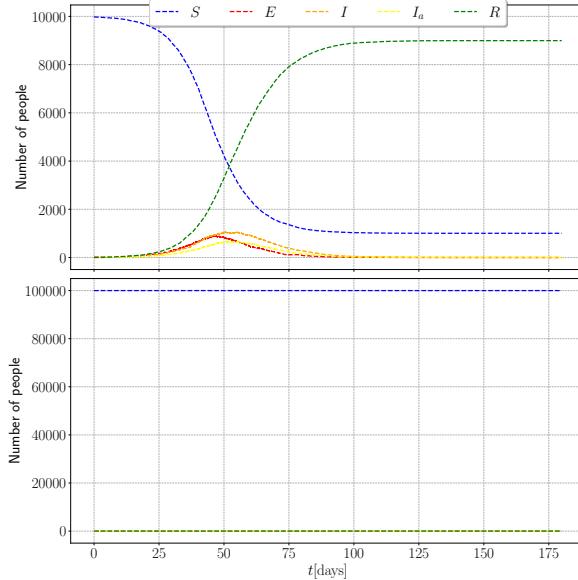


Figure 13: Solutions of Stochastic SEIIaR commuter model for the 2-city case, with the matrix  $\widetilde{\mathbf{M}}$  in equation (17).

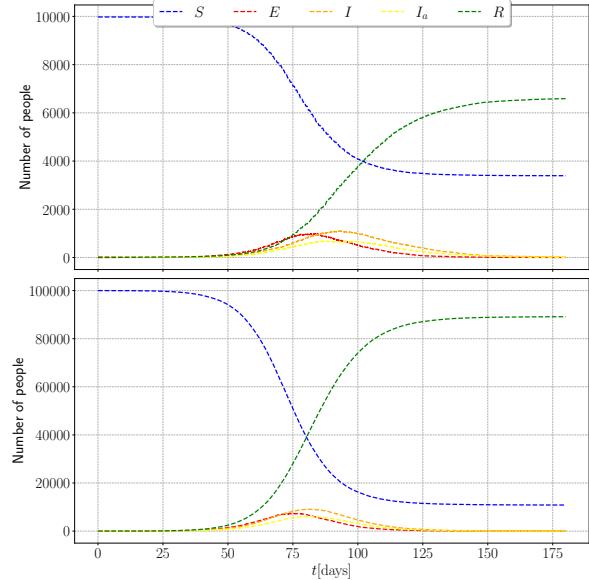


Figure 14: Solutions of Stochastic SEIIaR commuter model for the 2-city case, with the matrix  $\overline{\mathbf{M}}$  in equation (18).

Another quick check I made, which was even more crucial when dealing with the larger matrices in the proceeding problems, was to check that the number of people in each town are the same for each time step:

```
1 assert( np.all(np.sum(X, axis = (2,-1)) == np.sum(M, axis = 1)) ).
```

I also used this test in modifying the commuter solver for problem 2Eb, to check that it was consistent with this requirement.

## 5 Problem 2D: Larger stochastic SEIIaR Commuter model

### 5.1 a) 10 city simulation

We use the framework developed in the previous section to simulate a larger system of towns, now with 10 of them. The population matrix for this system is given in equation (11) in the problem sheet [Nor21]. We initialise the system with all people susceptible, except for 25 exposed in town 2. The time evolution of 10 realisations of the different states is shown in figure 15.

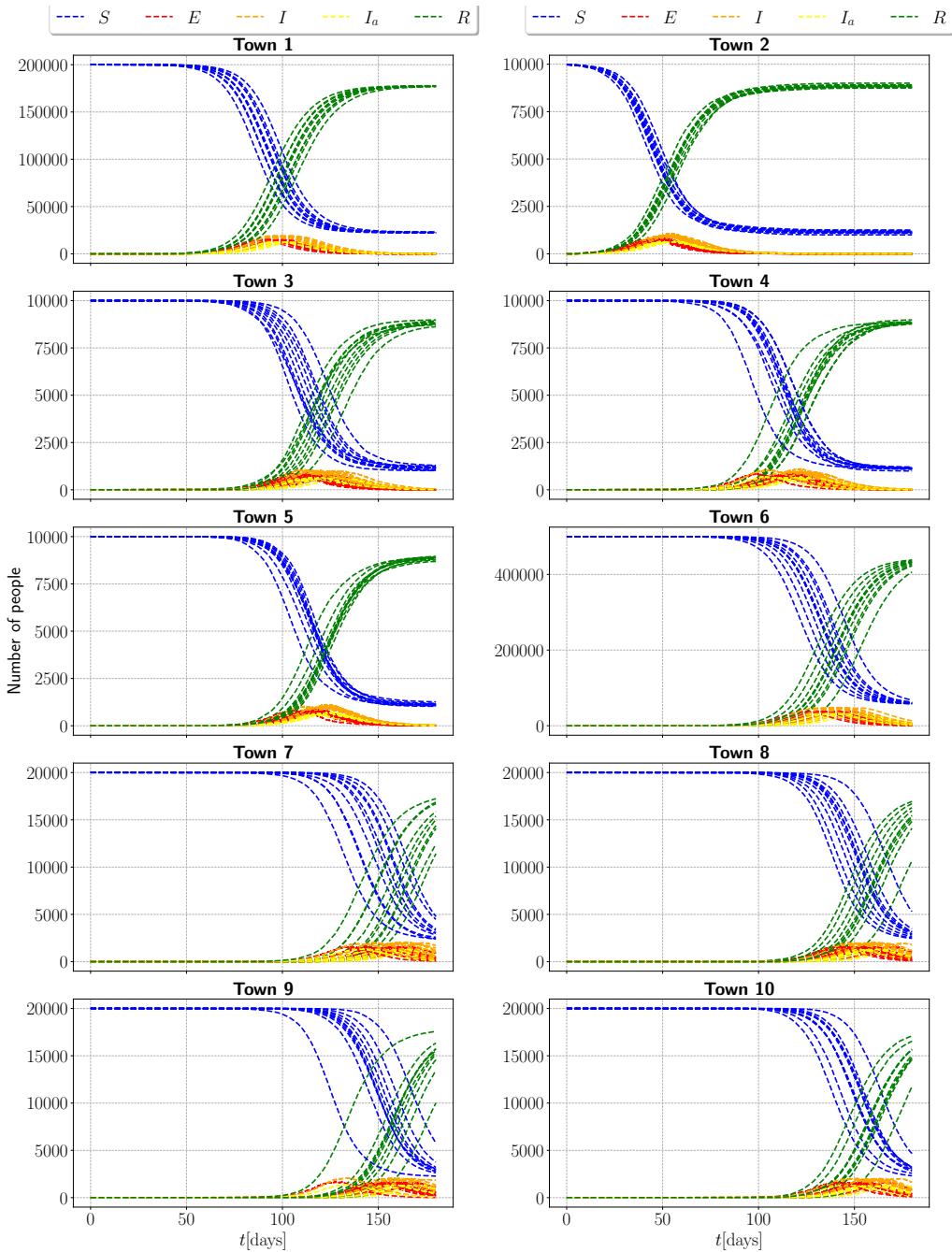


Figure 15: Solutions of Stochastic SEIIaR commuter model for the 10-city scenario.

Figure 15 clearly shows that the epidemic evolves fastest in town 2, where it began. This is as expected. Furthermore, we see that the second fastest evolution is in town 1, which is the closest connection to town 2 in the sense that the commuters of town 2 only travel to town 1. This is also a reassuring fact, indicating a correct implementation. Interestingly, for the towns with less connections to town 2 — e.g. town 9 and 10 — we see that the evolution lags approximately 100 days behind, and there seems to be a wider spread between each of the 10 realisations. The increasing spread may be explained by the fact that small delays in each realisation in the beginning become exceedingly large for another town, as some time must naturally pass for the infections to be exported here.

## 5.2 b) 356 city simulation

In this problem we use the full population structure handed out along with the problem set. We are here only interested in the number of municipalities with more than 10 infected people as a function of time ( $=: \mathcal{N}(t)$ ), so we modify the commuter solver described in algorithm 5 to only keep the current and previous state of the system at each time step, and to calculate  $\mathcal{N}$  for each time step, as described in the code overview section.

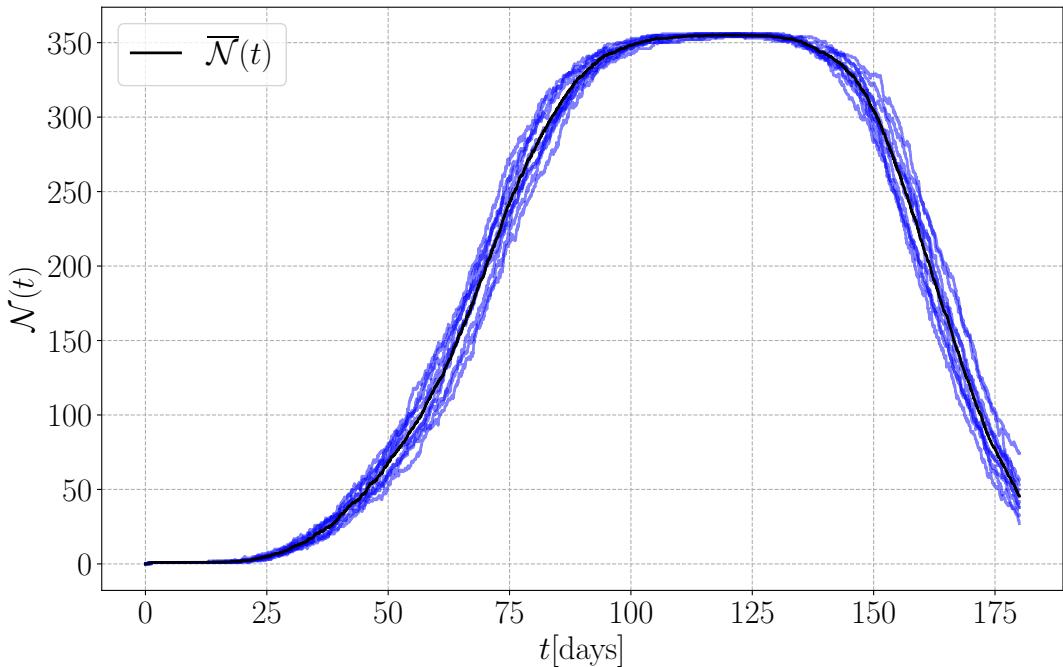


Figure 16: Number of municipalities with more than 10 infected people as a function of time. The mean is shown in the thick black line, while the 10 realisations are shown in opaque blue lines.

We initialise the system with 50 exposed people in municipality 1, and the rest being susceptible, and simulate for 180 days with the parameters specified in the problem sheet [Nor21].  $\mathcal{N}(t)$  for 10 realisations of this simulation is shown in figure 16. This figure shows that the infections have spread to all of the municipalities after roughly 110 days, and that a significant amount of people start to recover after 150 days. Only about 50 municipalities have more than 10 infected individuals when 180 days has passed, on average.

### 5.3 c) Reduced number of commuters in 356 city simulation

To study the effect of making more people work from home, we change the population matrix by dividing the off-diagonal elements by 10, rounding to the nearest integer, and adding the remaining number of people to the diagonal. We show visually the difference between these two population matrices in figure 17.

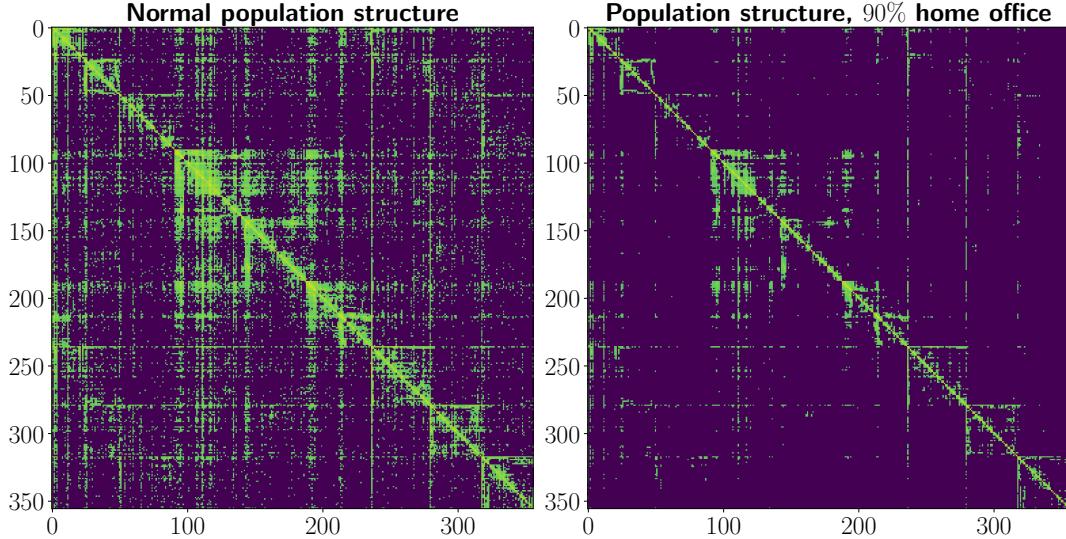


Figure 17: Population structures used for problem 2Eb and 2Ec respectively. The purple values are 0, and the numbers values of the entries are logarithmically scaled.

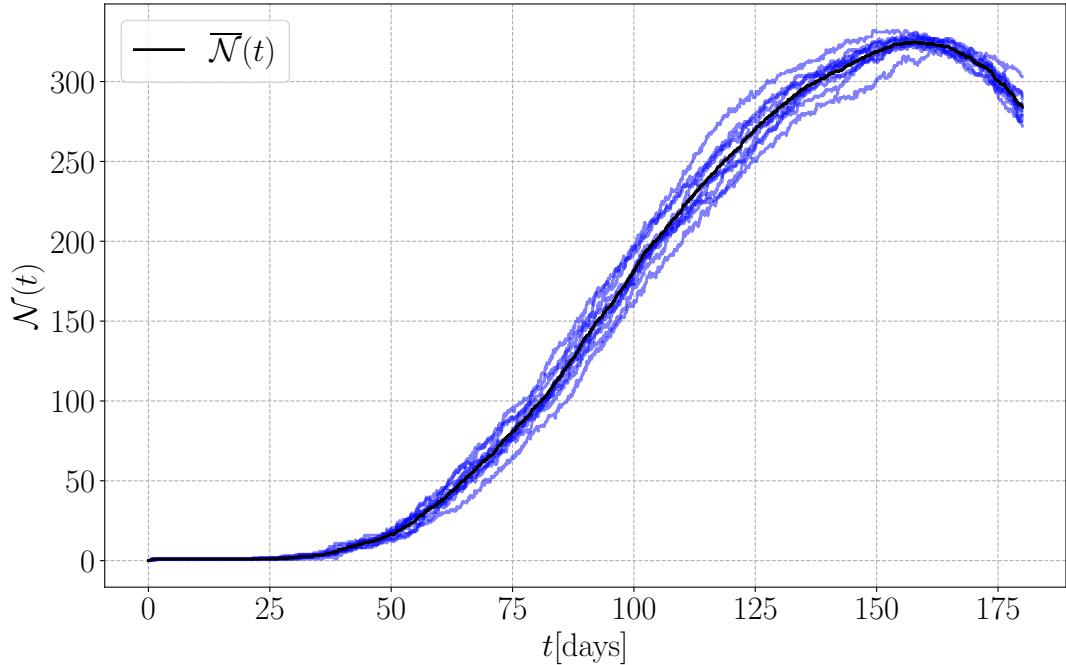


Figure 18: Number of municipalities with more than 10 infected people a a function of time, with reduced travelling. The mean is shown in the thick black line, while the 10 realisations are shown in opaque blue lines.

We run the same simulation as above, and display the number municipalities with more than 10 infected people as a function of time,  $\mathcal{N}(t)$ , in figure 18. Clearly, the effect of reduced travelling is to reduce the rate of infection spread. In figure 16 we see that  $\mathcal{N}$  saturates after about 110 days, while in figure 18 it never even reaches the case where  $\mathcal{N}$  is 356. The increase in  $\mathcal{N}(t)$  is seen to be much slower in this case, and the peak of the curve occurs after roughly 150 days. As  $\mathcal{N}(t)$  essentially is a measure of the degree of spread in the infections, we see that the effect of reducing travelling between municipalities reduces the geometric spread of the infections, exactly as we would intuitively expect. A side effect of reducing the spatial spreading of the infections is however to increase its spread in time. This is clearly seen by the fact that there are still over 200 municipalities with more than 10 infected people left after 180 days in the case of 90 % home office.

## References

- [Nor21] Tor Nordam. Exam 2021 - Computational physics TFY4235, 2021.
- [Was04] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer, New York, 2004.