NTNU
Institutt for Fysikk

Contact during the exam: Tor Nordam
tor.nordam@ntnu.no

Final exam 2021
TFY4235/FY8904
Computational Physics

Exam time: May 3, 09.00 - May 7, 16.00 (Norwegian time)

The answer to this exam should take the form of a report in pdf format. Additionally, you should submit your report from **Exercise 2** as a pdf or jupyter notebook (or similar), and the source code for both the exam and the exercise as a single zip file. Use filenames or folders to make it clear what source code belongs to the exam and the exercise. If you submit the exercise as a notebook, you do not need to attach the code for that separately. You may also include extra plots, animations, etc., in the zip file. These three files (exam, exercise, attachments) should be uploaded through the digital exam page on Inspera. If you for some reason have trouble using the upload function, you should as a backup solution send the reports and code to me by email.

Write your candidate number, not your name, on the reports. Collaboration is encouraged, so feel free to discuss the problems with anyone, but you should write both your code and your report yourself. Use of numerical libraries is of course permitted. You do not need to include all the theory from the assignment in your report, but you should write your report in such a way that someone with access to both the assignment and your report can easily follow and reproduce your work. As an example, you do not need to copy all the equations into your report, but referring to the equations in the problem set by number is helpful for the reader.

Some important points to keep in mind when you write your report:

- Describe what you have done in those cases where methods, parameters, etc. are left open for you to decide, and discuss your choices briefly.
- Describe any testing you have done, and what steps you have taken to convince yourself that your implementation is correct.
- Should you be unable to complete some tasks, describe what you have done. For example, if you have a mistake in your code that you are unable to find, describe the steps you have taken to find it. This will count positively, even if some tasks are left unfinished.

Any extra information during the exam (typos, clarifications, etc.) will be posted on Blackboard, so I recommend checking there from time to time.

Problem 1: Exercise 2
(counts 20% of the grade)

As problem 1, you should hand in your report from Exercise 2. Note that you are *not* required to submit animations in order to get full score on this exercise.

Problem 2: Epidemic modelling
(counts 80% of the grade)

This problem has five sub-problems, and each sub-problem counts approximately the same towards the final grade. They are intended to be completed in order.

In this problem, we will study some aspects of epidemic modelling, which is a topic you may have read about in the media lately. We will start by looking at a simple scheme called an SIR model. The model name comes from the three states it models: Susceptible, Infected, and Recovered. In this scheme, any infected person will eventually recover, and recovered people will not be re-infected. This is a very simplified model, but it can be used to address some qualitative questions. We will then proceed to expand the model, taking into account some of the additional features used by the Norwegian Institute of Public Health [1] in their modelling of the covid-19 situation.

Problem 2A: Deterministic SIR model

The simplest form of the SIR model is expressed as three coupled ordinary differential equations (ODEs). The variables are the number of susceptible people S, the number of infected people I, and the number of recovered people R, out of a total population of $N = S + I + R$. The ODEs describing the development are

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta \frac{IS}{N}, \tag{1a}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \beta \frac{IS}{N} - I/\tau, \tag{1b}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = I/\tau, \tag{1c}$$

where the parameter β describes the rate at which susceptible people encounter infected people and become infected, and τ is the typical duration of the infection.

Note that the parameter β describes the combined effect of two separate processes: The probability that people meet, and the probability that transmission of the infection occurs when people meet. Hence, the parameter β depends both on social patterns, and on the ease with which the particular disease we study is transmitted.

You may have heard about the so-called "R number" (not to be confused with the variable R in the SIR-model). In this model, we have $\mathcal{R}_0 = \beta\tau$, where $\mathcal{R}_0$ is the "R number", more properly called the basic reproduction number. It describes how many new infections each case is expected to produce, in a population where (almost) everyone is susceptible. Hence, if $\mathcal{R}_0 > 1$, the epidemic will grow exponentially, and if $\mathcal{R}_0 < 1$, the epidemic will die out. To find the exponential growth rate, we use that $\mathcal{R}_0$ is only valid when almost everyone are susceptible, which means that the fraction S/N will be very close to 1 (note that it cannot be larger than 1, since we must have $S \leq N$),

and hence the equation for the early development of I can be simplified to

$$\frac{\mathrm{d}I}{\mathrm{d}t} = (\beta - 1/\tau)I. \tag{2}$$

The model described here is a deterministic model: Given the same initial conditions and the same parameters, it will always give the same results. The model is also only valid for very large populations. The value of N has no particular meaning, and you may choose to set $N = 1$ and look at S, I, and R as the fractions of the population that are susceptible, infected and recovered.

In a case with initial values $R = 0$, $S/N = 1 - \epsilon$, and $I/N = \epsilon$, where ϵ is a small positive number, we can find semi-analytical expressions for the values of S and R as time goes to infinity (in a finite population, I will always go to zero as time goes to infinity). The expressions for $S(\infty)$ and $R(\infty)$ are [5]

$$S(\infty) = \exp\left[-\mathcal{R}_0\big(1 - S(\infty)\big)\right], \tag{3a}$$

$$R(\infty) = 1 - \exp\left[-\mathcal{R}_0 R(\infty)\right]. \tag{3b}$$

Eqs. (3) do not have regular solutions, but you can approximate them iteratively.

Tasks

a) Implement the model given by Eqs. (1). You may choose any ODE method, but comment on your choice. Let $\beta = 0.25\,\mathrm{day}^{-1}$, and $\tau = 10\,\mathrm{days}$. Assume that a fraction $I/N = 10^{-4}$ of the population is initially infected, and the rest is susceptible. Model the development for 180 days, and plot the results. Add the values of $S(\infty)$ and $R(\infty)$ from Eqs. (3) to your plot (see example in Fig. 1).

b) Plot the development of I as a function of time on a semi-log plot (logarithmic on the vertical axis), and confirm that the early phase of the development matches the solution of Eq. (2).

c) During the first months of the epidemic, there was discussion around the idea of "flattening the curve". The curve in question is the graph showing number of infected people as a function of time (see Fig. 1). The idea is that by reducing the number of social interactions, the value of β is reduced, and this leads to fewer people being infected simultaneously. Use your model to determine the largest value of β that will keep I/N smaller than 0.2 at the peak of the epidemic.

d) Another point that can be investigated with a simple model like this, is the question of how large a fraction of the population must be vaccinated before an outbreak becomes impossible. For simplicity, we will assume that vaccinated people are equivalent to recovered people: The do not become infected, and they do not infect others. Hence, you can simulate a (partially) vaccinated population by setting the value of R to something other than 0 at the start of the simulation. Use your model to determine the smallest vaccinated fraction that will prevent an outbreak from growing exponentially.

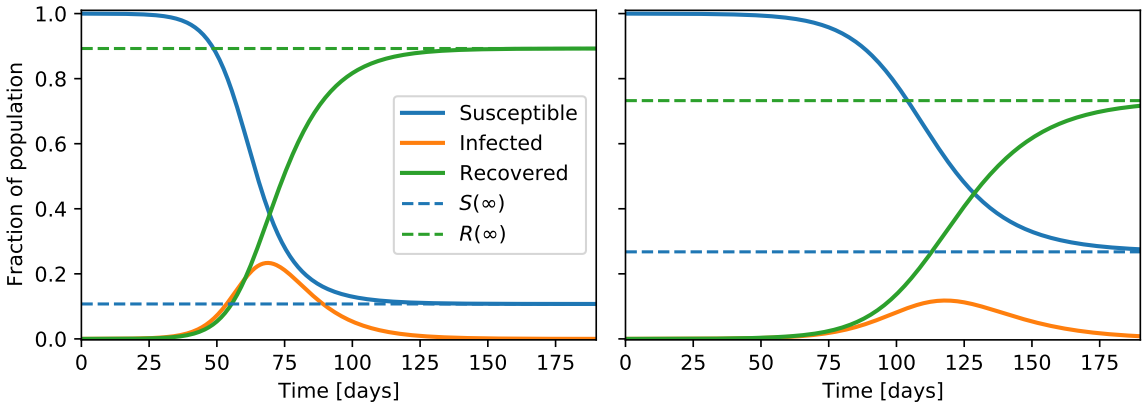

Figure 1: Flattening the curve: Reducing the parameter β has the effect of reducing the number of people who are infected simultaneously, potentially preventing the health services from being overwhelmed.

Problem 2B: Stochastic SIR model

Next, we will look at a stochastic version of the SIR model which is more applicable to smaller populations. The terms in this stochastic model are very similar to the ones in the deterministic model, but with two key differences:

- The numbers of people in the model are integers.
- Instead of rates, we have probabilities.

In this model, we have that S, I, and R are the *numbers* of people who are respectively susceptible, infected, and recovered, and the total population is $N = S + I + R$. During an interval Δt, each susceptible individual has some probability, $P_{S\to I}$, of being infected (if $I > 0$), and this probability is such that for a very large population the rate of infections is the same as in the deterministic model (with the same β and τ). Similarly, during an interval Δt, each infected person has some probability, $P_{I\to R}$, of recovering, and the probability is such that for a large population the rate of recovery becomes the same as in the deterministic model. The two probabilities $P_{S\to I}$ and $P_{I\to R}$ are given by

$$P_{S\to I} = 1 - \exp\left(-\Delta t \beta I/N\right), \tag{4a}$$
$$P_{I\to R} = 1 - \exp\left(-\Delta t/\tau\right). \tag{4b}$$

There are different ways of implementing such a stochastic model. The most practical approach here is probably to model the variables S, I, and R directly, so that we don't have to track every individual. At each timestep, we then have to draw random integers that represent the number of people transitioning between states. A set of discrete equations that describes how the variables change during an interval Δt is:

$$S(t+\Delta t) = S(t) - \Delta_{S\to I}, \tag{5a}$$
$$I(t+\Delta t) = I(t) + \Delta_{S\to I} - \Delta_{I\to R}, \tag{5b}$$
$$R(t+\Delta t) = R(t) + \Delta_{I\to R}, \tag{5c}$$

where

$$\Delta_{S\to I} = \mathcal{B}\big(S, P_{S\to I}\big), \tag{6a}$$
$$\Delta_{I\to R} = \mathcal{B}\big(I, P_{I\to R}\big). \tag{6b}$$

Here, $\mathcal{B}(n,p)$ is a random number from a binomial distribution, with n trials and probability of success p. See the Appendix for some additional comments.

Tasks

a) Implement a stochastic SIR model as described above. Let $\beta = 0.25\,\text{day}^{-1}$, and $\tau = 10\,\text{days}$. Set $N = 100\,000$, and assume that $I = 10$ people are initially infected. Model the development of the epidemic for 180 days, doing 10 simulation runs. Plot the results together with the results from the deterministic model for the same values of the parameters β and τ, and the same fraction of initially infected people. Discuss your choice of timestep for the two different models.

b) Plot the development of I for each of the 10 simulations as a function of time on a semi-log plot (logarithmic on the vertical axis), and confirm that the early phase of the development matches the solution of Eq. (2).

c) In a deterministic model where everyone is susceptible and where $\mathcal{R}_0 > 1$, any amount of infection will inevitably lead to an outbreak. However, in a stochastic model, a small outbreak may disappear by itself purely by chance. For a population $N = 100\,000$, and parameters $\beta = 0.25\,\text{day}^{-1}$, and $\tau = 10\,\text{days}$, use your model to numerically find the probability that an outbreak will disappear by itself, for initial number of infected individuals ranging from 1 to 10, and present the results in a plot. Comment on points like how you determine if an outbreak disappeared, the number of simulations you ran, and the uncertainty of your results.

Problem 2C: Stochastic SEIIaR model

In this subproblem, we will add another layer of realism to our model, based on the model that is in use by the Norwegian Institute of Public Health (FHI). You may find it useful to refer to their webpages [1], and there is also a description of such a model in one of their papers [4] (see the section called "The disease dynamics model" on page 8 in the reference). However, note that the model we use here differs a little from both the paper and the FHI webpages, so while it may be useful to refer to the paper and webpages for background, you should implement the model as described here.

The idea is that we add two extra states that people can be in, to provide a more realistic description. The states in this model are

- S, Susceptible
- E, Exposed (but not yet able to infect others)
- I, Infected
- I_a, Infected asymptomatic
- R, Recovered

People who are susceptible may become exposed if they meet someone who is infected. People who are exposed will become infected after some time, with some people showing symptoms of the disease, and some people not showing symptoms (asymptomatic). People who are infected will recover after some time.

The probabilities of the different transitions happening during an interval Δt are

$$P_{S\to E} = 1 - \exp\left(-\Delta t \beta \frac{r_s I + r_a I_a}{N}\right), \tag{7a}$$

$$P_{E\to I} = f_s \times \big(1 - \exp\left(-\Delta t/\tau_E\right)\big), \tag{7b}$$

$$P_{E\to I_a} = f_a \times \big(1 - \exp\left(-\Delta t/\tau_E\right)\big), \tag{7c}$$

$$P_{I\to R} = 1 - \exp\left(-\Delta t/\tau_I\right), \tag{7d}$$

$$P_{I_a\to R} = 1 - \exp\left(-\Delta t/\tau_I\right). \tag{7e}$$

The different parameters in Eqs. (7) are explained in Table 1.

When implementing Eqs. (7) into your stochastic SEIIaR model, the most practical approach is probably to use the multinomial distribution to implement the branching into two different categories of infected, I and I_a. Say that you have E exposed persons, and during an interval Δt each of those E persons will do exactly one of the following:

- Transition to infected (I), with probability $P_{E\to I}$
- Transition to infected asymptomatic (I_a), with probability $P_{E\to I_a}$
- Remain exposed (E), with probability $1 - (P_{E\to I_a} + P_{E\to I})$

Table 1: The parameters in the stochastic SEIIaR model.

Parameter	Value	Units	Description
β	0.55	day^{-1}	Transmission rate
r_s	1.0	-	Infectiousness when symptomatic
r_a	0.1	-	Infectiousness when asymptomatic
f_s	0.6	-	Symptomatic fraction
f_a	0.4	-	Asymptomatic fraction
τ_E	3	days	Typical time from exposure to infection
τ_I	7	days	Typical duration of infection

The numbers that respectively become infected, infected asymptomatic or remain exposed can then be modelled by random variables from a multinomial distribution:

$$\Delta_{E\to I}, \Delta_{E\to I_a}, \Delta_{E\to E} = \mathcal{M}\big(E, (P_{E\to I}, P_{E\to I_a}, 1 - P_{E\to I} - P_{E\to I_a})\big), \tag{8}$$

where we will always have that

$$\Delta_{E\to I} + \Delta_{E\to I_a} + \Delta_{E\to E} = E. \tag{9}$$

See also some additional comments on the multinomial distribution in the Appendix.

Tasks

a) Implement a stochastic SEIIaR model as described above, with the transition probabilities given by Eqs. (7), and the parameters given in Table 1. Set $N = 100\,000$, and assume that $E = 25$ people are initially exposed, and everyone else are susceptible. Model the development of the epidemic for 180 days, doing 10 simulation runs. Plot the time-development of the different variables. Add to the plot the solution of the deterministic model from problem 2A, with parameters $\beta = 0.25\,\text{day}^{-1}$ and $\tau = 10\,\text{days}$, and initial infected fraction $I/N = 10^{-4}$. Comment on the similarities and differences between the results from the deterministic SIR model and the stochastic SEIIaR model.

b) Among the measures that can be taken against the spread of a disease is to instruct people with symptoms to avoid contact with others. If enough people with symptoms self-isolate, it might be possible to eliminate any outbreaks. In our model, this can be implemented by setting the parameter r_s to some number smaller than 1 (see Table 1). Consider a case with $N = 100\,000$, where everyone is susceptible, except for $E = 25$ people who are initially exposed. Use your model to find how small r_s must be to get a situation where outbreaks do not grow exponentially. Comment on your approach, and the uncertainty in your results.

Problem 2D: Stochastic SEIIaR commuter model

In this subproblem, we will add the effects of geographical distribution and travel to our model. We will use a simple scheme, similar to the one used by FHI [1, 4]. Everybody in this model has a defined home town. This is where they spend the night and evening. Additionally, people have a town in which they spend the day. For most people, this will be the same town, but some people travel to another town during the day. The idea is to capture the effect of people traveling to another town during the day, for work, or school or other reasons. In this simplified model, we only consider these day-time travels, and each person who travels during the day has one fixed town that they travel to. This is called a commuter model [2].

For simplicity, we split each 24-hour period ("døgn", in Norwegian) into two 12-hour parts, that we call "day" and "night". During the night, people make up a part of

the population in their home town, and we run the stochastic SEIIaR model separately for each town with that population. During the day, people make up a part of the population of the town where they spend the day, and we run the stochastic SEIIaR model separately for each town with that population. The following description is taken from the webpages of the Norwegian Institute of Public Health [3, 2]:

> *"This model is a stochastic SEIIaR (susceptible, exposed, infectious, infectious asymptomatic, recovered) metapopulation model that including commuting. Each location has a local infection system, while the locations are connected by people who commute each day. The model differentiates between day and night. During the day you can infect/be infected in the location where you work, while during the night you can infect/be infected in the location where you live. It is the same commuters who travel back and forth each day. At the start of a day, all commuters are sent to their work location, where they mix for 12 hours. The commuters are then sent to their respective home locations, where they mix for 12 hours."*

We will describe the population structure by a matrix, where element i, j gives the number of people who live in Town i and work in Town j. Hence, the sum over the entries in row i gives the night-time population of Town i, while the sum over column j gives the day-time population of Town j.

As an example, the matrix

$$\begin{bmatrix} 9000 & 1000 \\ 200 & 99800 \end{bmatrix} \tag{10}$$

describes a distribution with two towns, called Town 1 and Town 2, where Town 1 has a population of 10 000, of which 9000 people stay in Town 1 during the day, while 1000 people travel to work in Town 2 during the day. Town 2 has a population of 100 000, of which 200 travel to work in Town 1 during the day, while the rest stay in Town 2. Hence this describes a fairly normal situation where some fraction of the population of a small town travels into a neighbouring larger town for work during the day, while relatively few people travel in the other direction.

An important point is that the people who travel during the daytime are not selected at random from the population of a town. Rather, it is the same people that travel every day. This is a more realistic representation of travel patterns, and selecting people to travel at random would give different dynamics in the transmission of the disease.

There are different approaches you can take to implementing the stochastic SEIIaR commuter model. You will somehow need to keep track of the number of people in the different states (S, E, I, Ia, R), for each group of people (i.e., each element in the population matrix). While the matrix in Eq. (10) is small enough that you can probably write a "custom" code for this problem, you might want to think about how to make a more general implementation that can handle any matrix. This will be useful for problem 2E.

Tasks

a) Set up a population structure as described by the matrix in Eq. (10). Assume that 25 people in Town 1 are initially exposed (E), and all the other people are susceptible (S). Implement and run the stochastic SEIIaR commuter model as described above, with the parameters given in Table 1. Plot the development of the five states, separately for the population of each town (i.e., for all the people that *live* in that town), and comment on the results.

b) Describe your implementation, and comment on any steps you took to test your code. For example, did you run your model for any special cases of the population distribution matrix, where you can reason about the expected behaviour?

Problem 2E: Larger stochastic SEIIaR commuter simulations

In this subproblem, you will apply the stochastic SEIIaR commuter model developed above to two different larger population distributions. The first population structure is given by the matrix below, and illustrated schematically in Fig. 2. It describes a semi-realistic population structure of two larger towns, each surrounded by four smaller towns.

$$\begin{bmatrix} 198600 & 100 & 100 & 100 & 100 & 1000 & 0 & 0 & 0 & 0 \\ 500 & 9500 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 500 & 0 & 9500 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 500 & 0 & 0 & 9500 & 0 & 0 & 0 & 0 & 0 & 0 \\ 500 & 0 & 0 & 0 & 9500 & 0 & 0 & 0 & 0 & 0 \\ 1000 & 0 & 0 & 0 & 0 & 498200 & 200 & 200 & 200 & 200 \\ 0 & 0 & 0 & 0 & 0 & 1000 & 19000 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1000 & 0 & 19000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1000 & 0 & 0 & 19000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1000 & 0 & 0 & 0 & 19000 \end{bmatrix} \tag{11}$$

On Blackboard, along with the exam problem, you will also find a text file called `population_structure.csv` which contains a square matrix of size 356×356. This matrix describes the distribution of the working population of Norway[1]: Element i, j contains the number of people who live in municipality i, and work in municipality j. The matrix is stored on a csv (comma-separated values) format. The file contains 356 lines, and each line has 356 entries, separated by commas. Each entry is an integer number of people. The structure of the matrix is illustrated in Fig. 3.

The idea is still the same, but at a larger scale: People are at the location where they work during the daytime, and at the location where they live during the nighttime. This is of course a simplified model, but it gives us the opportunity to study the spread of an epidemic in a country when daily travel patterns are taken into account.

[1]The data are sourced from Statistics Norway (SSB): `https://www.ssb.no/statbank/table/03321`

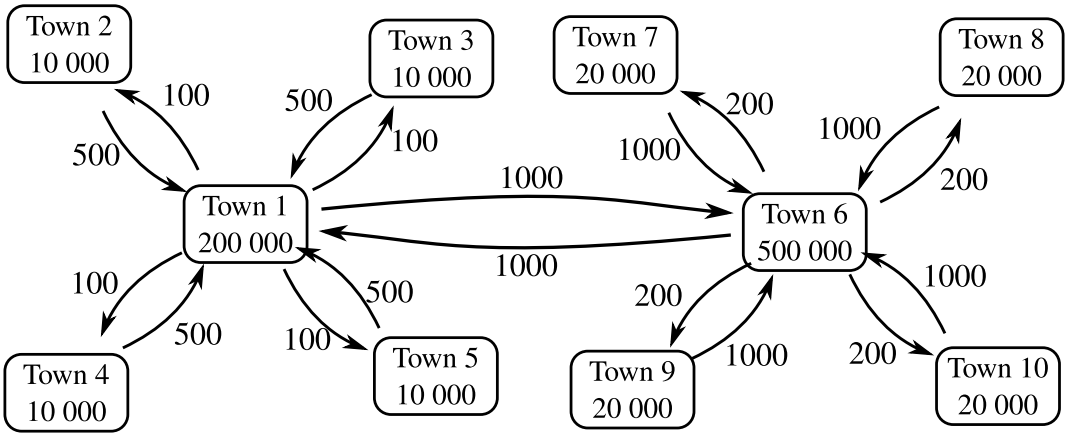

Figure 2: A schematic illustration of the connections between the towns in the population structure described by the matrix in Eq. (11). The numbers in the boxes give the night-time populations of each town, and the arrows indicate the number of people who travel for work during the day.

Tasks

a) Set up a population structure as described by the matrix in Eq. (11). Assume that 25 people in Town 2 are initially exposed (E), and all others are susceptible (S). Run the stochastic SEIIaR commuter model for 180 days, with the parameters given in Table 1. Do 10 simulation runs. Plot the time-development of the five states (S, E, I, I_a, R), separately for the population of each town (i.e., the people that *live* in that town). Comment on the results.

b) Load the population structure matrix from the file `population_structure.csv`, which you will find on Blackboard. Run your stochastic SEIIaR commuter model for 180 days on this population structure, with the parameters given in Table 1. Assume that everyone is initially susceptible, except for $E = 50$ exposed people who both live and work in the first municipality (Oslo) in the matrix. Do 10 simulation runs. Plot the number of municipalities with more than 10 infected people ($I+I_a > 10$, among the people who *live* in that municipality), as a function of time, for all 10 simulation runs.

c) To study the effect of reduced travel, you will modify the population distribution as follows: For each row in the population distribution matrix, you will divide the off-diagonal elements by 10 (and round to the nearest integer). This will reduce the number of people that travel for work by 90%. You will then have to add a corresponding number of people to the diagonal element on that row, to keep the number of people that live in each municipality unchanged. With this change, 90% the people who formerly traveled for work will now work from home instead. Repeat the 10 simulations from the previous task but with the modified population structure, and again plot the number of municipalities with more than 10 infected people, as a function of time.

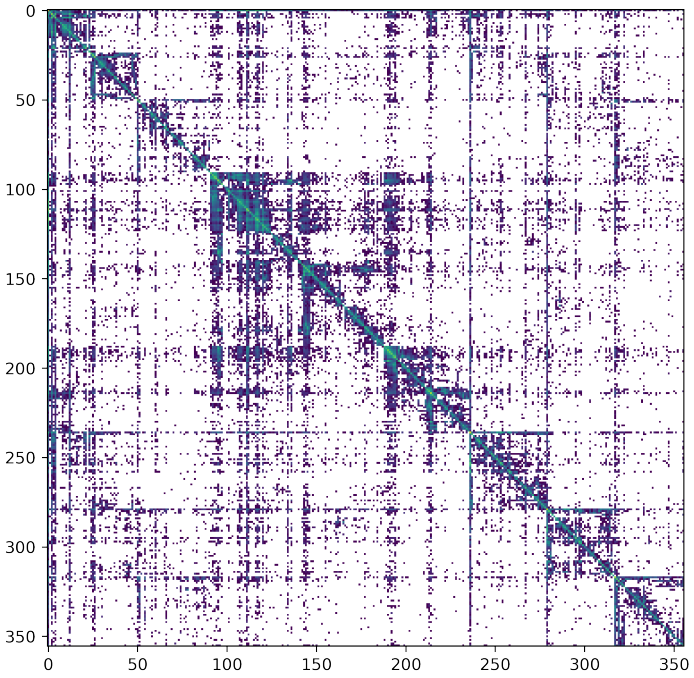

Figure 3: Distribution of the working population of Norway. Element i, j is the number of people living in municipality i and working in municipality j. The values are log-scaled and white cells are 0.

Appendix

Binomial distribution

Think of an experiment that consists of checking the outcome of some random process a number of times. For example, throw a six-sided die $n = 100$ times, and count the number of sixes. At each throw, the probability of a six is $p = 1/6$. If you do the experiment several times, you will usually get a different number of sixes each time. The distribution of the number of sixes is drawn from the binomial distribution $\mathcal{B}(100, 1/6)$. More generally, the number of successes in an experiment with n trials, and probability of success p, is a random integer from the binomial distribution $\mathcal{B}(n, p)$.

Mulitnomial distribution

The multinomial distribution is a generalisation of the binomial distribution, useful in cases with several *mutually exclusive* outcomes. For example, say you throw a six-sided die $n = 100$ times, and you count the number of times you get a 6, the number of times you get an odd number (1, 3 or 5), and the number of times you get something else (2 or 4). These events are mutually exclusive, and their probabilities are respectively 1/6, 1/2 and 1/3. If you repeat the experiment you will usually get a different outcome, and the three numbers will be random integers from a multinomial distribution $\mathcal{M}\big(100, (1/6, 1/2, 1/3)\big)$. If you use an implementation of a multinomial distribution from a library, check the documentation to confirm that it behaves as you expect.

References

[1] https://www.fhi.no/en/id/infectious-diseases/coronavirus/coronavirus-modelling-at-the-niph-fhi/.

[2] https://folkehelseinstituttet.github.io/spread/articles/commuter_model.html.

[3] https://folkehelseinstituttet.github.io/spread/.

[4] Solveig Engebretsen, Kenth Engø-Monsen, Arnoldo Frigessi, and Birgitte Freiesleben de Blasio. A theoretical single-parameter model for urbanisation to study infectious disease spread and interventions. *PLoS computational biology*, 15(3), 2019. https://doi.org/10.1371/journal.pcbi.1006879.

[5] Joel C Miller. A note on the derivation of epidemic final sizes. *Bulletin of mathematical biology*, 74(9):2125–2141, 2012. https://doi.org/10.1007/s11538-012-9749-6.