Project 1 TMA4265

Eirik Fagerbakke and Henrik Grenersen

2022-10-09

Problem 1

a)

An outbreak of measles can be modeled as a Markov Chain.

The reason why it can be modeled as a Markov chain, is because it follows the Markov property, which entails that the probabilities of the current step are only affected by the previous step. A person who has been susceptible for 10 days is for instance just as likely to be infected as a person who has been susceptible for just 1 day.

We can model the Markov chain with the following probability matrix:

$$P = \begin{bmatrix} 1-\beta & \beta & 0 \\ 0 & 1-\gamma & \gamma \\ \alpha & 0 & 1-\alpha \end{bmatrix}$$

It is impossible to go from susceptible to recovered (you need to get infected first), from infected to susceptible (you need to be recovered first), and from recovered to infected (you need to be susceptible first). These probabilities are therefore all 0.

The probability of going from recovered to susceptible is α, the probability of going from susceptible to infected is β, and the probability of going from infected to recovered is γ.

The rest of the probabilities are found by the fact that each row needs to sum to 1.

b)

With $\alpha = 0.005$, $\beta = 0.01$ and $\gamma = 0.1$, we get

$$P = \begin{bmatrix} 0.99 & 0.01 & 0 \\ 0 & 0.9 & 0.1 \\ 0.005 & 0 & 0.995 \end{bmatrix}$$

To show that the chain actually has a limiting distribution, we can determine if it is regular. If so, it also has such a distribution. To determine this, we first consider P^2

$$P^2 = \begin{bmatrix} 0.09 & 0.01 & 0 \\ 0 & 0.90 & 0.10 \\ 0.005 & 0 & 0.995 \end{bmatrix} \begin{bmatrix} 0.09 & 0.01 & 0 \\ 0 & 0.90 & 0.10 \\ 0.005 & 0 & 0.995 \end{bmatrix} = \begin{bmatrix} 0.0081 & 0.0099 & 0.001 \\ 0.0005 & 0.81 & 0.1895 \\ 0.005425 & 0.00005 & 0.990025 \end{bmatrix}$$

Since all the elements are larger than 0, P is regular and a limiting distribution exists.

Now, we are interested in finding the limiting distribution, as it can be used to calculate the long-run mean number of days spent in each state. To find the limiting distribution, we solve

$$\pi_0 = 0.99\pi_0 + 0.005\pi_2$$

$$\pi_2 = 0.1\pi_1 + 0.995\pi_2$$

$$\pi_0 + \pi_1 + \pi_2 = 1$$

Solving the first equation for π_0 :

$$\pi_0 = \frac{0.005}{0.01}\pi_2$$

Solving the second equation for π_1 :

$$\pi_1 = \frac{0.005}{0.1}\pi_2$$

Plugging this into the third equation:

$$\frac{0.005}{0.01}\pi_2 + \frac{0.005}{0.1}\pi_2 + \pi_2 = 1 \quad \Rightarrow \quad \pi_2 = \frac{20}{31} \quad \pi_0 = \frac{0.005}{0.01}\frac{20}{31} = \frac{10}{31} \quad \pi_1 = \frac{0.005}{0.1}\frac{20}{31} = \frac{1}{31}$$

Using the limiting distribution, we can determine the mean number of days spent being in the different states for an individual. This relation is given by $365\pi_i$. We then get that the mean number of days spent being susceptible in a year is 117.7 days, being infected is 11.77 and being recovered is 235.5

c)

From the simulation, we obtain that in a year, we spend 150.5 days being susceptible, 6.4 days being infected and 208.1 being recovered. This is very similar to the results obtained in b).

To calculate the confidence interval, we need to look at the student T-distribution, as we do not know the actual mean and standard deviation. We instead need to calculate estimates for these based on the simulations. To do this, we have used t.test():

```
susceptibleConfInt <- t.test(susceptibleVec, conf.level = 0.95)$conf.int
infectedConfInt <- t.test(infectedVec, conf.level = 0.95)$conf.int
recoveredConfInt <- t.test(recoveredVec, conf.level = 0.95)$conf.int
```

	Lower bound	Upper bound	Theoretical value
Number of susceptible days in a year	105.7	129.1	117.7
Number of infected days in a year	10.75	13.64	11.77
Number of recovered days in a year	223.4	247.4	235.5

We see that these results are compatible with our results in b), as the theoretical values lie in all the confidence intervals. In a 95 % confidence interval, there is a 95 % chance that the theoretical values is contained in each interval in each simulation.

d)

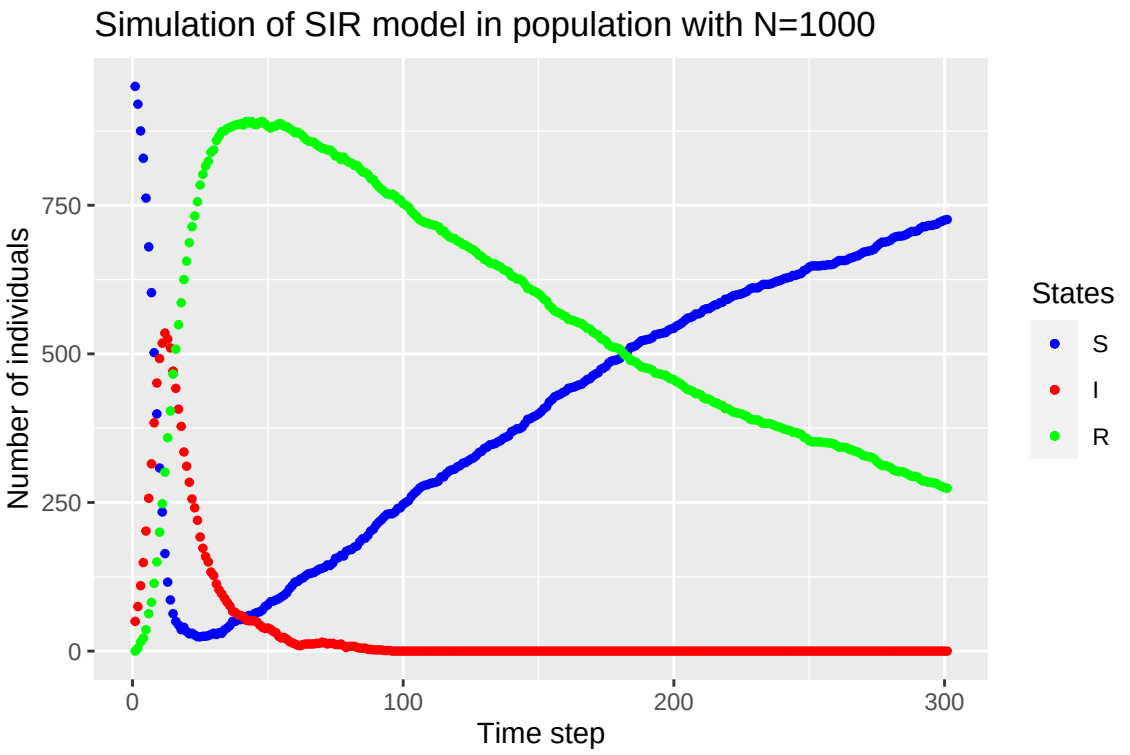
Since β changes with each time step, the probabilities are no longer stationary, but the processes might still be Markov if they fulfill the Markov property. Since α and γ are constant, we only need to calculate the value of β for the current time step, which is given by $\beta_n = \frac{I_n}{2N}$.

The process I_n, where we only know the number of infected individuals, is not a Markov process. Since we only know the number of infected individuals, we could look at the values of I_n in previous time steps, as we could use this to give an estimate of the number of susceptible and recovering individuals. By using information from multiple previous time steps, the process no longer fulfills the Markov property.

When adding S_n, we also effectively know the number of recovered individuals, as we know the population size. Z_n is therefore a Markov chain, as we only need to know the values of Z_n in the current time step to find probabilities for the next step. Adding the number of recovered individuals, R_n, does not cancel out the Markov property either, and therefore Y_n is also a Markov process.

Further, we can model each step as a binomial distribution. If we want to calculate S_n, we need to look at S_{n-1} and R_{n-1}, as these individuals could become susceptible again in the current step. We then have a binomial distribution with S_{n-1} individuals and a probability $1-\beta$ for becoming susceptible, and a binomial distribution with R_{n-1} individuals and a probability α for becoming susceptible. S_n will then be the sum of "successes".

e)



We see that for the time steps 0-50, we have a drastic increase in the number of infected people, and with this a drastic decrease in the number of susceptible people. When the number of infected people reaches its peak, the probability to become infected is also at its peak, but the number of susceptible individuals is so low that the number of infected individuals has to decrease.

We also see a sharp increase in the number of recovered individuals, as there are a lot of infected people who can become recovered. The probability that one stays recovered is very high (0.995), and for the time steps 50-300 we therefore see a stabilization.

f)

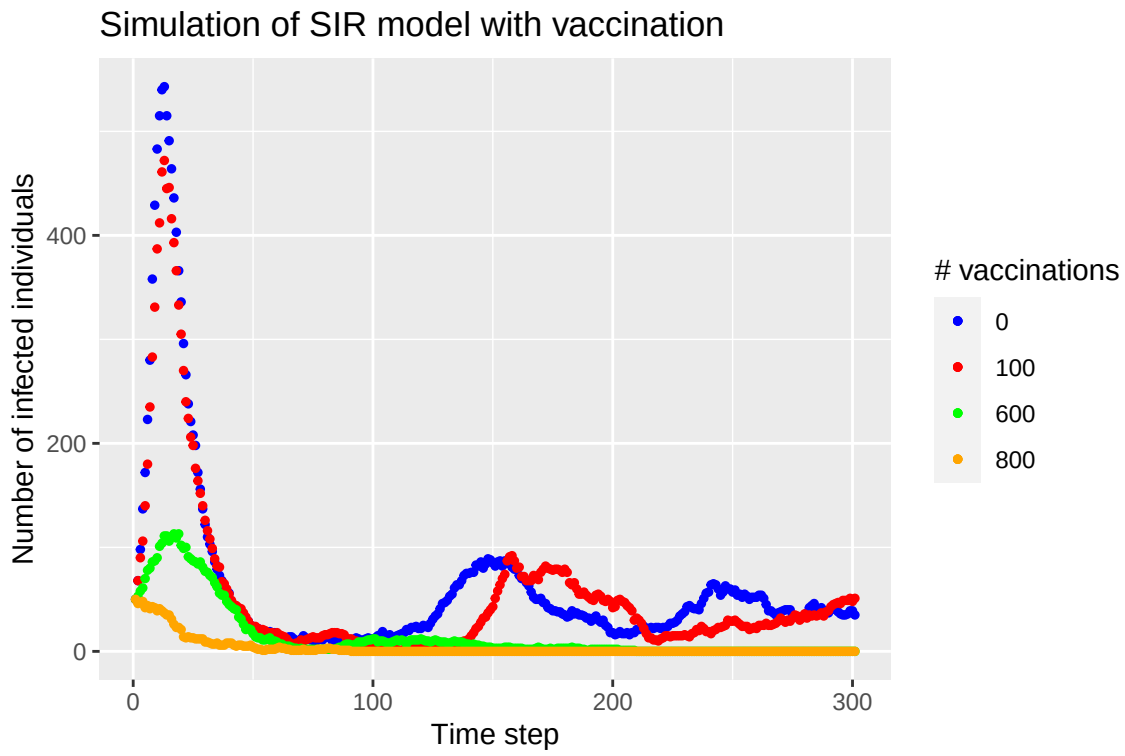
The computed confidence intervals for the maximum number of infected individuals and the corresponding time steps are computed below

	Lower bound	Upper bound	Expected value
max $\{I_n\}$	519.9	522.6	521.2
Time step	12.82	12.93	12.88

We see that the maximum number of individuals is reached very quickly, and that this is consistent with all the simulations, as the confidence interval is small. This maximum number is also rather high, as more than half of the population would have been infected at the same time, which could be very severe.

g)

We now want to see how a vaccination program could affect the outbreak. We will look at four different scenarios: 0, 100, 600 and 800 vaccinated individuals. We will assume that the individuals become immune after vaccination, and that they no longer will be able to pass on the disease. At the same time, we have chosen to use the whole population size for the probability of becoming infected $\beta_n = \frac{I_n}{2N}$. This makes sense if we assume that the vaccinations are distributed equally throughout the population. This would then mean that if we had a population of size 1000, with 998 vaccinated individuals, where one is infected and one is susceptible, the probability of becoming infected in the next time step would not be $\frac{1}{4}$, but rather $\frac{1}{2000}$.



		Lower bound	Upper bound	Expected value
0 Vaccinations	max $\{I_n\}$	522.6	525.1	523.9
	Time step	12.78	12.89	12.83
100 Vaccinations	max $\{I_n\}$	436.9	439.4	438.2
	Time step	13.57	13.69	13.63
600 Vaccinations	max $\{I_n\}$	97.08	98.64	97.86
	Time step	16.43	16.94	16.68
800 Vaccinations	max $\{I_n\}$	51.17	51.44	51.31
	Time step	1.921	2.145	2.033

Problem 2

a)

The probability of getting more than 100 claims in 59 days, is

$$\Pr\{X(59) > 100\} = \Pr\{X(59) - X(0) > 100\} = 1 - \Pr\{X(59) - X(0) \leq 100\}$$

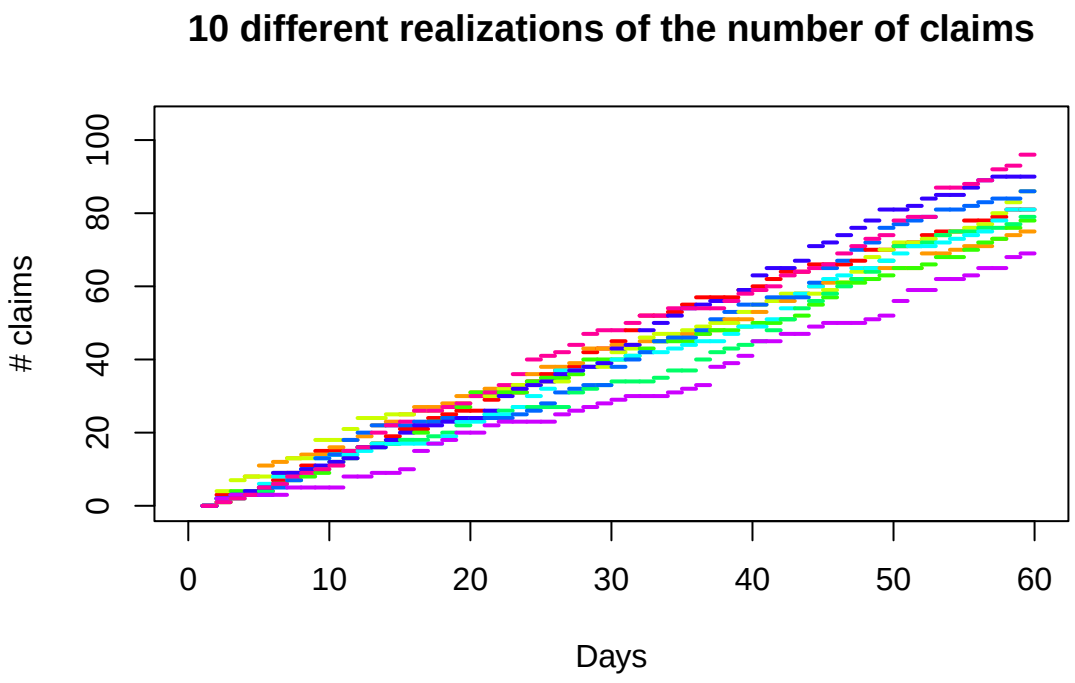
Since the interval has length 59, we need to calculate the cumulative distribution of a poisson process with rate $\lambda t = 1.5 \cdot 59 = 88.5$ with $x = 100$

$$1 - \Pr\{X(59) - X(0) \leq 100\} = 1 - \sum_{i=0}^{100} \Pr\{X(59) - X(0) = i\} = 1 - \sum_{i=0}^{100} \frac{88.5^i}{i!} e^{-88.5}$$

Calculating this in R, we obtain

$$\Pr\{\text{X(59)>100}\} = 0.10282$$

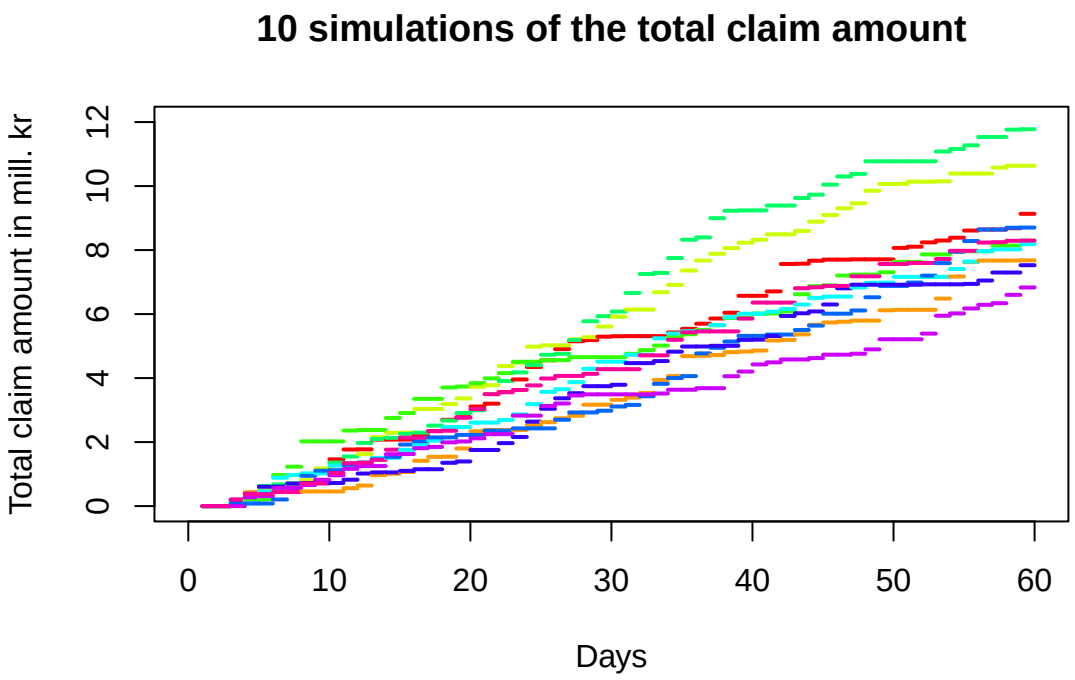
Now we want to simulate this process, and present 10 of these realizations in the plot below.



Now we have estimated the probability to be 0.089 from our 1000 simulations. This is not far from the theoretical value we calculated earlier, with an absolute difference of only 0.013822

b)

We will now look into the sum of the claim amounts reached in $t = 59$ days. That is $Z(t) = \sum_{i=1}^{X(t)} C_i$ with $C_i \sim \text{Exp}(\gamma)$ in mill. kr and $\gamma = 10$.



Here we have estimated the probability of the sum of the claim amounts exceeding 8 million kr in 59 days to be 0.704 from 1000 simulations.

c)

Now we can think about the variable $Y(t)$, which is the number of claims up to time t which have exceeded 250000 kr. We then need to take into account the number of claims received up to time t, $X(t)$, which is of course Poisson distributed as earlier, with rate λt.

$$Y(t) = \sum_{i=1}^{X(t)} I_i$$

Where each I is Bernoulli-distributed, with probability p of success, that is, that a specific claim exceeds 250000 kr. Now if we fix $X(t) = x$, we could then condition our $Y(t)$ on N as follows

$$Y(t)|X(t) = x = \begin{cases} 0, & x = 0 \\ \sum_{i=1}^{x} I_i & x > 0 \end{cases}$$

Now it is clear that this variable is binomially distributed, so

$$(Y(t)|X(t) = x) \sim \text{Binomial(x, p)}$$

With marginal distribution $p(Y|x) = \binom{x}{y} p^x (1-x)^{x-y}$. If we now calculate the marginal distribution using the law of total probability, we get that

$$\begin{aligned}
\Pr\{Y = y\} &= \sum_{x=0}^{\infty} \Pr\{Y = y, X = x\} = 0 + \sum_{x=y}^{\infty} \Pr\{Y = y, X = x\} \\
&= \sum_{x=0}^{\infty} \Pr\{X = x\} \Pr\{Y = y | X = x\} \\
&= \sum_{x=y}^{\infty} \frac{(\lambda t)^x e^{-\lambda t}}{x!} \binom{x}{y} p^y (1-p)^{x-y} \\
&= \sum_{x=y}^{\infty} \frac{(\lambda t)^x e^{-\lambda t}}{x!} \frac{x!}{y!(x-y)!} p^y (1-p)^{x-y} \\
&= \frac{e^{-\lambda t} p^y}{y!} (\lambda t)^y \sum_{x=y}^{\infty} \frac{(\lambda t (1-p))^{x-y}}{(x-y)!} \\
&= \frac{e^{-\lambda t} (\lambda p t)^y}{y!} \cdot e^{\lambda t (1-p)} \\
&= \frac{(\lambda p t)^y}{y!} e^{-\lambda p t} \\
&\implies Y(t) \sim \text{Poisson}(\lambda p t)
\end{aligned}$$

Now, the final part consists of determining p, that is, the probability that a single claim exceeds 250000 kr, and should therefore be examined. Remembering that each claim amount, C_i, is exponentially distributed, with rate $\gamma = 10$, we get that

$$p = \Pr\{C_i > 0.25\} = 1 - \Pr\{C_i \leq 0.25\} = 1 - \int_0^{0.25} \gamma e^{-\gamma c} \text{dc} = 1 + [e^{-\gamma c}]_0^{0.25} = e^{-0.25 \cdot \gamma}$$

Which all together gives that

$$Y(t) \sim \text{Poisson}(1.5 e^{-2.5} t) \quad \square$$