Project 1 : SIR-model and Insurance Claims

Kristian Willassen, Marc Eric Halsvik Müller

Problem 1

Problem 1.a)

We have the transition probability matrix

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

$\mathbf{P}$ is a Markov-chain because it has no memory. The parameters α, β and γ have all the information about the next step, and does not depend on previous steps, only on the current state. Thus, the matrix $\mathbf{P}$ upholds the Markov property.

Problem 1.b)

With values $\beta = 0.01$, $\gamma = 0.10$ and $\alpha = 0.005$ we get the matrix

$$\mathbf{P} = \begin{pmatrix} 0.99 & 0.01 & 0 \\ 0 & 0.90 & 0.10 \\ 0.005 & 0 & 0.995 \end{pmatrix}$$

If $\mathbf{P}^{(n)}$ has only positive nonzero elements for some n, in other words are regular, we now that the states of $\mathbf{P}$ has a limiting distribution π. We can check the first iteratives of $\mathbf{P}^n$, $n = 1, 2, 3$ to see if the matrix-elements tends toward a nonzero positive real number for all $\mathbf{P}^n_{ij}$.

$$\mathbf{P^2} = \begin{pmatrix} 0.9801 & 0.0189 & 0.001 \\ 0.0005 & 0.81 & 0.1895 \\ 0.009925 & 0.00005 & 0.990025 \end{pmatrix}$$

$$\mathbf{P^3} = \begin{pmatrix} 0.970304 & 0.026811 & 0.002885 \\ 0.0014425 & 0.729005 & 0.2695525 \\ 0.014775875 & 0.00014425 & 0.985079875 \end{pmatrix}$$

We can see that the $\mathbf{P}^3$ has only nonzero positive elements, which means that it is regular and therefore has a limiting distribution π.

Since π_i is the fraction of time spent in state i, we can find the number of days spent in state i by multiplying π_i by the number of days in a year.

For each component of π we then have

$$\pi_i = \sum_{k=0}^{N} \pi_k \mathbf{P}_{ki} \tag{1}$$

which can again be used to make the linear equations

$$\begin{aligned}
\pi_0 &= \pi_0\mathbf{P}_{00} + \pi_1\mathbf{P}_{10} + \pi_2\mathbf{P}_{20} = 0.99\pi_0 + 0.005\pi_2 \\
\pi_1 &= \pi_0\mathbf{P}_{01} + \pi_1\mathbf{P}_{11} + \pi_2\mathbf{P}_{21} = 0.01\pi_0 + 0.90\pi_1 \\
\pi_2 &= \pi_0\mathbf{P}_{02} + \pi_1\mathbf{P}_{12} + \pi_2\mathbf{P}_{22} = 0.10\pi_1 + 0.995\pi_2
\end{aligned}$$

We list the equations we shall use:

$$\begin{array}{cc}
I & -0.01\pi_0 + 0\pi_1 + 0.005\pi_2 = 0 \\
II & 0.01\pi_0 - 0.10\pi_1 + 0\pi_2 = 0 \\
III & \pi_0 + \pi_1 + \pi_2 = 1
\end{array}$$

Solving this system we get π componentwise:

$$\begin{aligned}
\pi_0 &= \quad 0.32258 \\
\pi_1 &= \quad 0.03226 \\
\pi_2 &= \quad 0.64516
\end{aligned}$$

We can multiply these values by the amount of days in a year (365) which gives us the number of days spent in each state on average:

State 0	(Susceptible) :	117.74	days
State 1	(Infected) :	11.77	days
State 2	(Recovered) :	235.49	days

Problem 1.c)

Using our simulator to run the simulation for 7300 timesteps we estimate the long-running means. The estimates can be found in table 1. These estimates are based on one run of the simulation. For more representative numbers we find confidence intervals instead.

Description	S	I	R
Absolute # (10 years)	1097	128	2425
# per year	109.7	12.8	242.5
Percentage	30%	4%	66%

Table 1: Estimates of long-running means

We calculate our 95% confidence intervals by simulating from the same initial state 30 times and then computing the confidence intervals from the results of these simulations. We have here chosen to compute 95% confidence intervals. These 95% confidence intervals can be found in table 2.

State (in code)	# days in state
Susceptible (0)	119.475 ± 14.345
Infected (1)	11.535 ± 1.435
Recovered (2)	233.985 ± 14.735

Table 2: 95% confidence intervals for the long running means.

Comparing these confidence intervals to the caculated limiting distribution π, we can see that for each state they give corresponding values for the amount of each state per year. The calculated limiting distributions are also located within our 95% confidence intervals for all states, supporting this point.

Problem 1.d)

I_n lacks the required information to be memoryless, thus not upholding the Markov-property. The reason being that the number of infected I_n are dependent on S_n and R_n, which means that if you only have I_n you can not implicitly know the vaule of the rest of the variables by just looking at the current state. We show this more formally:

$$P(I_{n+1} = i) = P(I_{n+1} = i | I_n = a, R_n = b, S_n = c)$$

but since we do not have R_n or S_n we must define their probability as

$$\begin{aligned} P(R_n = b) = & \quad P(R_n = b | R_{n-1} = x, S_{n-1} = y, I_{n-1} = z) \\ P(S_n = c) = & \quad P(S_n = c | R_{n-1} = x, S_{n-1} = y, I_{n-1} = z) \end{aligned}$$

which shows that we are dependent on the previous steps and not only the current step.

For Z_n we do not have all the information we need to uphold the Markov-property directly, but we do have all the information implicitly. $Z_n = (I_n, S_n)$, and thus since we know the total number of people N we can define $R_n = N - (S_n + I_n)$ which gives us

$$P(Z_{n+1} = q) = P(Z_n = d | I_n = i, S_n = s, (N - I_n - S_n) = r)$$

We can see that we here have all the information needed to be dependent only on the current state to predict the next state.

The same argument holds for Y_n, only here we have all the information directly, and thus need not redefine any of the arguments to uphold the Markov-property.

■

Problem 1.e)

We can split the simulation into the start phase, from the start until approximately timestep 50, and the long-run phase. In the first phase of the simulation there is a rapid rise in the number of infected individuals. This rise creates a spike that disappears as more and more individuals transition to the recovered state.

In the long-run phase the number of individuals in the different states are more stable. There are less transitions to the infected state, which again reduces the chance of individuals to be infected. Towards timestep 150-200 in our simulation there is a second explosive wave of infected, but as there are many recovered individuals, who have no probability of transitioning to infected, the peak of this wave is much lower than in the start phase.

Figure 1: Temporal evolution where starting number of infected, I_0, is equal to 50.

Problem 1.f)

To assess the potential severity of an outbreak we can use our simulator to predict how the outbreak will evolve in time. We do this by running the simulator with the same initial state many times. In our case we have chosen to use 1000 simulations. We can then find confidence intervals for when the outbreak will peak, and how much of the popualtion will be infected at the peak. The severity of the outbreak can then be assessed from when the peak is expected to happen and how large the portion of the population it affects. Our calculated confidence intervals of the number of infected individuals and timestep of this peak can be found in table 3.

max # infected	time of max # infected
524.473 ± 1.305	11.872 ± 0.051

Table 3: Number and timestep of max # infected. These are given as 95% confidence intervals.

Problem 1.g)

To account for vaccinated individuals we add a fourth state to our model, representing these individuals. This expands our transition probability matrix $\mathbf{P}$. We define our new $\mathbf{P}$ as

$$\mathbf{P} = \begin{pmatrix} 1-\frac{0.5I_n}{N} & \frac{0.5I_n}{N} & 0 & 0 \\ 0 & 1-\gamma & \gamma & 0 \\ \alpha & 0 & 1-\alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

where we also replace our parameter β with the parameter $\frac{0.5 \cdot I_n}{N}$ which makes an individuals chance of infection dependant on the the ratio of infected individuals in the population.

When adding the vaccinated-state to our model and updating our transition probability matrix we assume that the probability of an individual to transition to any other state is equal to 0. For this project we take this as a reasonable assumption, as we only deal with short time intervals. We now run the simulation again for different ratios of vaccination. In each scenario we start with 50 infected. This way we can compare how an outbreak of the same magnitude evolves differently in populations with different ratios of vaccination. We plot these four simulations in figure 2.

From figure 2 we can see how the four scenarios behave differently. As the portion of the popluation who are vaccinated rises the probability of susceptible individuals being infected will be reduced. We find this from the probability of being infected

$$\mathbf{P}_{01} = 0.5 \cdot \frac{I_n}{N}$$

As vaccinated individuals still contribute to the total population, N, while being inaccessible from the susceptible, infected and recovered states the chance of infection will be lower.

Problem 2

Problem 2.a)

We have that

$$X(t) \sim Poisson(\lambda t) = Poisson(1.5t)$$

# vaccinated	max # infected	time of max # infected
0 from **(f)**	524.473 ± 1.305	11.872 ± 0.051
100	439.025 ± 1.242	12.683 ± 0.061
600	97.079 ± 0.794	16.029 ± 0.280
800	51.273 ± 0.125	0.995 ± 0.106

Table 4: Number and timestep of max # infected when a portion of the population are vaccinated. The numbers given are 95% confidence intervals based on 1000 simulations of each scenario.

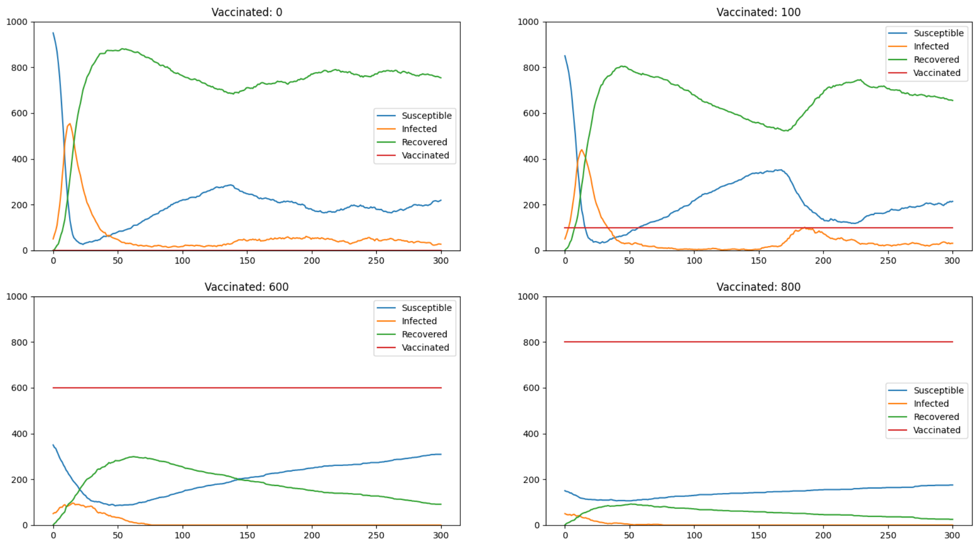

Figure 2: SIRV model, with differing amounts of vaccinated individuals.

and we want to find

$$P(X(59) > 100)$$

so we use the cumulative distribution for the poisson-process and get

$$P(X(59) > 100) = 1 - \sum_{s=0}^{100} \frac{(1.5 \cdot 59)^s}{s} e^{-1.5 \cdot 59} = 0.1028$$

The computed average was 0.1031. The relative error between these corresponding values is 0.29% which confirms that our calculations correspond. The realization-values vary for the most part between 0.12 and 0.80.

For a figure that shows 10 realizations of $X(t), 0 \leq t \leq 59$ we refer to "Insurance Claim Amounts" in Figure 2.

Problem 2.b

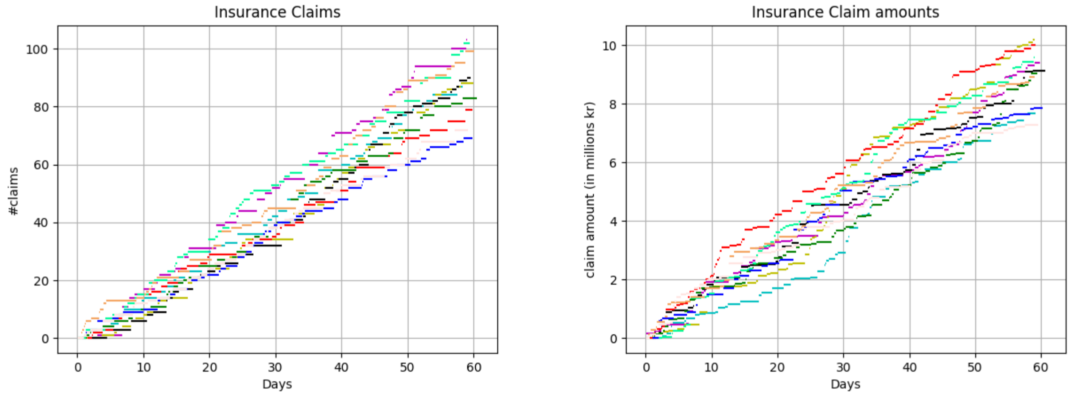

Figure 3: (Left) Shows 10 different realizations of number of claims per day.
(Right) Shows 10 differents realizations of claim-amounts per day (in millions)

We calculated the estimated probability to be 0.719. This means that the majority of the realizations in "Insurance Claim Amount" in Figure 3 should end up above 8, which it does. This reinforces our result.

Problem 2.c

We prove that $\{Y(t) : t > 0\}$ is a poisson-process by using the law of total probability:

$$Pr\{Y(t) = y\} = \sum_{x=y}^{\infty} Pr\{Y(t) = y|X(t) = x\}Pr\{X(t) = x\} = \sum_{x=y}^{\infty} P^y(1-p)^{x-y} \frac{x!}{y!(x-y)!} \frac{(\lambda t)^x}{x!} e^{-\lambda t} =$$

$$\frac{(\lambda Pt)^y}{y!} e^{-\lambda t} \sum_{x=y}^{\infty} \frac{(\lambda t(1-P))^{x-y}}{(x-y)!} = \frac{(\lambda Pt)^y}{y!} e^{-\lambda t} \cdot e^{\lambda t(1-P)} = \frac{(\lambda Pt)^y}{y!} e^{-\lambda Pt}$$

This shows that $Y(t) \sim Poisson(\lambda Pt)$ with rate-parameter $\lambda P = 1.5P$. We have that $C_i = Exp(\gamma)$ for $\gamma = 10$, and then use the cumulative exponential distribution to find the value of P:

$$P = 1 - \int_0^{0.25} 1 - e^{-\gamma x} dx = 0.082$$

We then get by

$$rate = \lambda P = 1.5 \cdot 0.082 = 0.123$$

that the rate-parameter is 0.123.