Project 1

Group members:
Ola Runeson Rasmussen
Sigurd Kampevold Johanson
Hanne Blomdal Laupstad

October 5, 2022

Problem 1

a)

For this task, we consider only one individual, and we let X_n denote the individual's state at time n. We also assume $X_0 = 0$. In short $X_n : n = 0, 1, ...$ is a Markov chain because it is a stochastic process satisfying the Markov property, i.e., the next state does only depend on the present state, and the rest of the past does not affect it. This could be formalized as the following:

$$Pr\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0\} = Pr\{X_{n+1} = j | X_n = i\}$$

for $n = 0, 1, ...$, and for all states i and j.

To put it in a more problem-specific way, whether the individual is infected tomorrow does only depend on which condition it is in today, etc.

The problem states that if an individual is susceptible today, it has a probability β of being infected tomorrow, else it will remain susceptible, i.e., it has a probability $1 - \beta$ of remaining susceptible. If it is infected today, it has a probability γ of being recovered tomorrow, else it will stay infected with a probability $1 - \gamma$. If it is recovered, and therefore immune, today, it has a probability γ of losing its immunity and becoming susceptible tomorrow, and it has thus a probability of $1 - \alpha$ of staying immune tomorrow as well. As S, I and R corresponds to the states 0, 1 and 2 respectively, these transition dynamics give raise to the following transition matrix $\mathbf{P}$:

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

b)

With the given values for the parameters α, β and γ, we get the following transition probability matrix:

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

From this, we see that the Markov chain consists of exactly one equivalence class as all states communicate. By this, we mean that it is possible to get from each state to all others. Consequently, the Markov chain is irreducible.Regarding periodicity, we can easily verify that the Markov chain is aperiodic as each state has self-loops. Thus, each state has period 1. Moreover, as we only have one finite and closed equivalence class, this has to be recurrent as of the lecture notes. This could also be easily verified, as the expected number of visits to each state must be infinite. For a finite state space, this does as well imply positive recurrence.

Due to the Markov chain satisfying positive recurrence, aperiodicity and irreducibility, the chain satisifies the condition for Theorem 4.4, and we can be certain that the chain has a limiting distribution. The limiting distribution can be found through solving the equations

$$\begin{cases} \sum_{i=0}^{3} \pi_i = 1 \\ \pi_j = \sum_{i=0}^{3} \pi_i \mathbf{P}_{ij} \ , \ j = 0, 1, 2 \end{cases}$$

which give raise to the equations

$$\begin{cases} \pi_0 + \pi_1 + \pi_2 = 1 \\ \pi_0 = 0.99\pi_0 + 0.005\pi_2 \\ \pi_1 = 0.01\pi_0 + 0.9\pi_1 \\ \pi_2 = 0.1\pi_2 + 0.995\pi_2 \end{cases}$$

These equations evaluate to

$$\begin{cases} \pi_0 = \frac{10}{31} \\ \pi_1 = \frac{1}{31} \\ \pi_2 = \frac{20}{31} \end{cases}$$

As the limiting distribution gives us the fraction of time spent in each state, we can find the expected number of days spent in each state by multiplying 365 with the limiting probabilities. Thus, we get

$$\mathbb{E}[\#\text{days being susceptible}] = \frac{10}{31} \times 365 = 117.74$$

$$\mathbb{E}[\#\text{days being infected}] = \frac{1}{31} \times 365 = 11.77$$

$$\mathbb{E}[\#\text{days being immune}] = \frac{20}{31} \times 365 = 235.48$$

c)

Please find the code related to this problem in the separate file.

To compute the 95% confidence intervals, we ran the code 30 times. For each iteration, we collected the estimated days spent in each state per year in three separate vectors. Afterwards, we calculated the 0.025 and 0.975 percentiles, giving us the lower and upper bounds for the confidence intervals respectively. The computed confidence intervals are the following:

$$\text{For days spent in state 0: } [77.18, 150.19]$$

$$\text{For days spent in state 1: } [6.79, 17.71]$$

$$\text{For days spent in state 2: } [200.38, 275.40]$$

We do see that the computed confidence intervals contain the computed estimates from **b)**, and we could therefore say that they are compatible. However, we should note that neither 7300 time steps nor 30 iterations are large numbers, probably keeping the intervals quite large.

d)

To determine whether the stochastic processes are Markov chains, we must discuss whether they satisfy the Markov property. For $\{I_n : n = 0, 1, 2, ...\}$, the present state indicates the number of infected individuals in the population. Thus, the state space is $\{0, 1, ..., 999, 1000\}$. The next state, i.e., the number of infected individuals tomorrow depend on the number of infected individuals today, i.e., the present state, but also on the number of susceptible individuals S_n and indirectly the number of recovered individuals R_n. Thus, the next state does not only depend on the current state, and the process does not satisfy the Markov property. Hence, it is not a Markov chain.

For $\{Y_n : n = 0, 1, 2, ...\}$, $Y_n = (S_n, I_n, R_n)$, each state represents the number of susceptible individuals, the number of infected individuals and the number of recovered individuals together. Thus, we see clearly that the next state only depends on the present state, i.e., the distribution between the three conditions tomorrow depends on the the distribution between them today, as long as the individuals are independent of each other. Thus, the process satisfy the Markov property and is a Markov chain. For $\{Z_n : n = 0, 1, 2, ...\}$, $Z_n = (I_n, S_n)$, each state represent the number of susceptible individuals and the number of infected individuals. Given that the population size is constant and known, the number of recovered individuals is known implicitly through the model as

$R_n = N - I_n - S_n$. By this, the same argument yields as for Y_n, and the stochastic process is a Markov chain. Hence, $\{Y_n\}$ and $\{Z_n\}$ are Markov chains, while $\{I_n\}$ is not.

e)

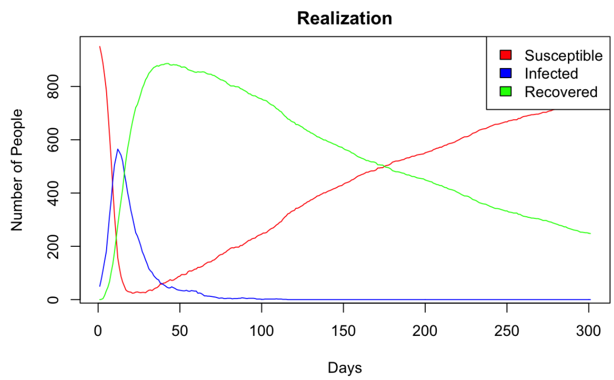

Figure 1: One realization of the temporal evolution for S_n, I_n and R_n

As of Figure 1, we see that the behaviour of the Markov chain is different when comparing the steps $0-50$ and $50-300$. Initially, a large proportion of the population is susceptible, where the next step is to get infected. As more and more people get infected, the probability of going from susceptible to infected increases due to the dynamics of the process. Thus, we see a spike in the number of infected people early on, whereas the number of susceptible drop rapidly. From then on, the number of recovered individuals grows, as the probability of recovering is rather large relative to the rest of the transitions. Thus, the number of infected decreases rapidly again, while the number of recovered reaches its maximum before time step 50. From time step 50 and onwards, the main trend in the evolution is that the number of susceptible grows while the number of recovered decreases. The number of infected seems to reach zero, making it impossible to get infected due to the dynamics of the system. Thus, being susceptible turns into an absorbing state for each individual.

f)

As of the attached code file, by utilizing the mean as estimator for the expectancy, we estimate the following values:

$$\mathbb{E}[\max\{I_0, I_1, ..., I_{300}\}] = 522.501$$

$$\mathbb{E}[\min\{\arg\max_n\{I_0, I_1, ..., I_{300}\}\}] = 12.842$$

As for the confidence intervals, the following were estimated:

$$\mathbb{E}[\max\{I_0, I_1, ..., I_{300}\}] : \ [481, 564]$$

$$\mathbb{E}[\min\{\arg\max_n\{I_0, I_1, ..., I_{300}\}\}] : \ [11, 14.025]$$

By utilizing the confidence intervals, we can assess the potential severity of the outbreak. In 95% of the cases, we will have the maximum number of infected people at the same time to be between 481 and 564. Thus, a bit more than the half of the population will be infected at the same time, something which may cause problems for the health institutions. Moreover, we see that we reach this top just a short while after the start of the outbreak, i.e., between 11 and 14 days after the start, given that 50 individuals were infected in the beginning. This implies a potentially rapid growth, implying that the society does not have a large time to prepare for a large part of the population being infected.

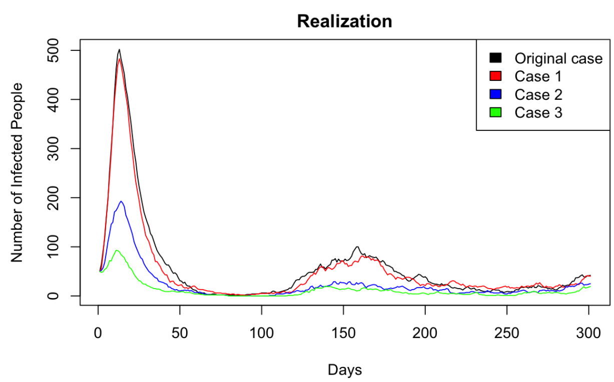

Figure 2: Temporal evolution for I_n in the different cases

g)

As of Figure 2, we see clear differences when introducing a number of vaccinated people. When only introducing 100 vaccinated people, there are small differences in the temporal evolution, as the estimated maximum number of infected individuals decrease from 522 to 459. The differences are clearer for case 2 and 3, where 600 and 800 vaccinated people are introduced respectively. This decreases the estimated maximum number of infected individuals 199 and 95 for the two cases, which is clearly shown in the temporal evolution. In these cases, we do not see the increasing number around 150 time steps as of the original case and case 1. Thus, introducing vaccination seems to have an effect for the temporal evolution and the potential severity of an outbreak. We do also note that the estimated time steps to reach its maximum number is within the computed confidence interval for all cases. Hence, it does not vary significantly when introducing vaccination.

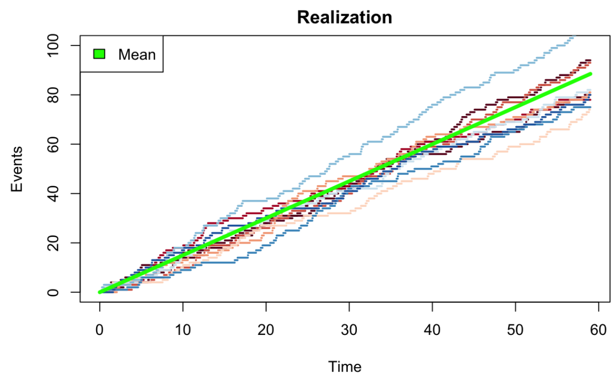

Figure 3: Simulations of $X(t)$

Problem 2

a)

For this problem we are calculating the probability that there are more than 100 claims about the insurance company before March 1st. By using the cumulative distribution for the Poisson process, we calculate the probability:

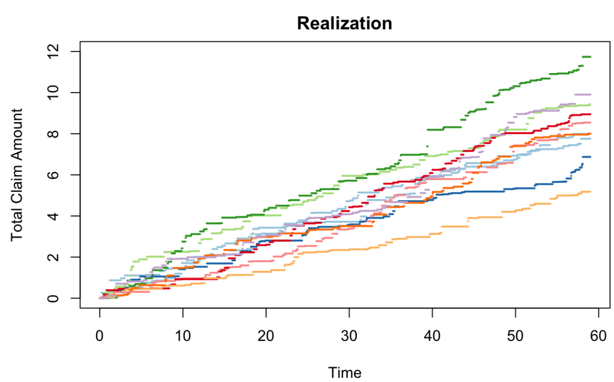

Figure 4: Simulations of $Z(t)$

c)

By the law of total probability, we prove that $\{Y(t) : t \geq 0\}$ is a Poisson process:

$$\begin{aligned}
Pr\{Y(t) = k\} &= \sum_{n=0}^{\infty} Pr\{Y(t) = k|X(t) = n\}Pr\{X(t) = n\} \\
&= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\frac{(\lambda t)^n e^{-\lambda t}}{n!} \\
&= \frac{e^{-\lambda t}(\lambda t p)^k}{k!}\sum_{n=k}^{\infty}\frac{[\lambda t(1-p)]^{n-k}}{(n-k)!} \\
&= \frac{e^{-\lambda t}(\lambda t p)^k}{k!}e^{\lambda t(1-p)} \\
&= \frac{e^{-\lambda t p}(\lambda t p)^k}{k!} \\
&\text{for } k = 0, 1, ...
\end{aligned}$$

$\{Y(t) : t \geq 0\}$ is therefore a Poisson process, with the rate λp. For $\gamma = 10$, we got $C_i \sim Exp(\gamma)$. We also have $\lambda = 1.5$, and have to find the value of p, that can be calculated by using the exponential distribution:

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

The rate is therefore calculated as: $\lambda p = 1.5 \cdot 0.082 = 0.123$