Project 1: Daniel Hostadt, Tiago Pereira

Problem 1: Modelling an outbreak of measles

a)

A (discrete-time) Markov chain is a discrete-time stochastic process $\{X_n : n = 0, 1, ...\}$ that satisfies the Markov property

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

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

The situation at hand $\{X_n : n = 0, 1, ...\}$ matches the definition of a Markov chain. It is a stochastic process and it satisfies the Markov property 1, i.e., if a person stays in its current state or if it moves to the next state only depends on the current state and not on past states.

The transition probabilities are given as follows:

- If a person is in state 0, it stays in state 0 with probability $1 - \beta$ and moves to state 1 with probability β.
- If a person is in state 1, it stays in state 1 with probability $1 - \gamma$ and moves to state 2 with probability γ.
- If a person is in state 2, it stays in state 2 with probability $1 - \alpha$ and moves to state 0 with probability α.

Thus, the transition probability matrix is given by:

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

b)

The transition probability matrix $\mathbf{P}$ is given by:

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

By matrix multiplication, the two-step transition probability matrix $\mathbf{P^2}$ is given by:

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

We see that the Markov chain is regular and has finite state space. Thus, by theorem 4.1, the limiting distribution $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2)$ exists and is the unique non-negative solution of the equations

$$\begin{aligned} \pi_j &= \sum_{k=0}^{N} \pi_k P_{kj}, \quad j = 0, 1, 2 \\ \sum_{k=0}^{N} \pi_k &= 1 \end{aligned} \tag{2}$$

Inserting yields:

$$\pi_0 = 0.99\pi_0 + 0.005\pi_2 \tag{3}$$
$$\pi_1 = 0.9\pi_1 + 0.01\pi_0 \tag{4}$$
$$\pi_2 = 0.995\pi_2 + 0.1\pi_1 \tag{5}$$
$$\pi_0 + \pi_1 + \pi_2 = 1 \tag{6}$$

Simplifying 3 and 4,

$$\pi_0 = 0.5\pi_2 \tag{7}$$
$$\pi_1 = 0.1\pi_0 \tag{8}$$

Inserting 7 into 8 and subsequent into 6,

$$\pi_1 = 0.05\pi_2 \tag{9}$$
$$0.5\pi_2 + 0.05\pi_2 + \pi_2 = 1 \tag{10}$$

Solving 10, 9, and 7,

$$\pi_2 = \frac{20}{31} \tag{11}$$
$$\pi_1 = \frac{1}{31} \tag{12}$$
$$\pi_0 = \frac{10}{31} \tag{13}$$

and $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2) = (\frac{10}{31}, \frac{1}{31}, \frac{20}{31})$.

The long-run mean number of days per year spent in each state is the limiting distribution multiplied by the number of days in a year. Thus,

$$(\frac{10}{31}, \frac{1}{31}, \frac{20}{31}) * 365 = (\frac{3650}{31}, \frac{365}{31}, \frac{7300}{31}) \approx (117.7, 11.8, 235.5) \tag{14}$$

c)

As stated in the `R code`, the 95% confidence intervals are

$$\pi_0 \in [108.0, 130.6] \tag{15}$$
$$\pi_1 \in [10.5, 13.6] \tag{16}$$
$$\pi_2 \in [221.7, 245.6] \tag{17}$$

These CIs are compatible with the exact solutions of subtask b) since the exact values are an element of the CI for each state. CIs were constructed as follows: We used the `R` function `t.test()` and applied it on the vector of 30 values obtained from 30 simulations. The function assumes the data to be approximately normally distributed and is based on the Central Limit Theorem. Additionally, it uses a t-distribution since the sample size it comparably small with only 30 values.

d)

Recalling the definition of a Markov chain from subtask a), $\{I_n : n = 0, 1, ...\}$ is a Markov chain since the state of the stochastic process in the next period only depends on the state of the process in the current period. The same holds for the processes $\{Z_n : n = 0, 1, ...\}$ and $\{Y_n : n = 0, 1, ...\}$. Recall that the state space is $\{S_n \in \{0, 1, ..., 1000\}, I_n \in \{0, 1, ..., 1000\}, R_n \in \{0, 1, ..., 1000\}, S_n + I_n + R_n = 1000\}$. How many people belong to which group of S_n, I_n, R_n in the next period only depends on how many people belong to which group in the current period. The transition probability matrix nonetheless is considerably large but finite since there are many possible combinations.

e)

Figure 1 shows one evolution of the number of peoples in the groups S, I, and R. The behaviour in the interval 50-300 is more or less stable since the Markov chain has a limiting distribution and the process varies around that distribution. The behaviour in the interval 0-50 is different because the initial distribution is very skewed compared to the limiting distribution and the adjustment process takes its time.

f)

Based on 1000 simulations the 95% confidence intervals are:

- $E[max\{I_0, I_1, ..., I_{300}\}] \in [522.7, 525.3]$
- $E[min\{argmax_{n \leq 300}\{I_n\}\}] \in [12.803, 12.903]$

Both measurements are indicators of the severity of an outbreak: The first one tells about how many people are expected to being infected in the peak of the outbreak. The second one tells about how long it takes to reach the peak of the outbreak if initially no one is recovered and 50 people are infected.

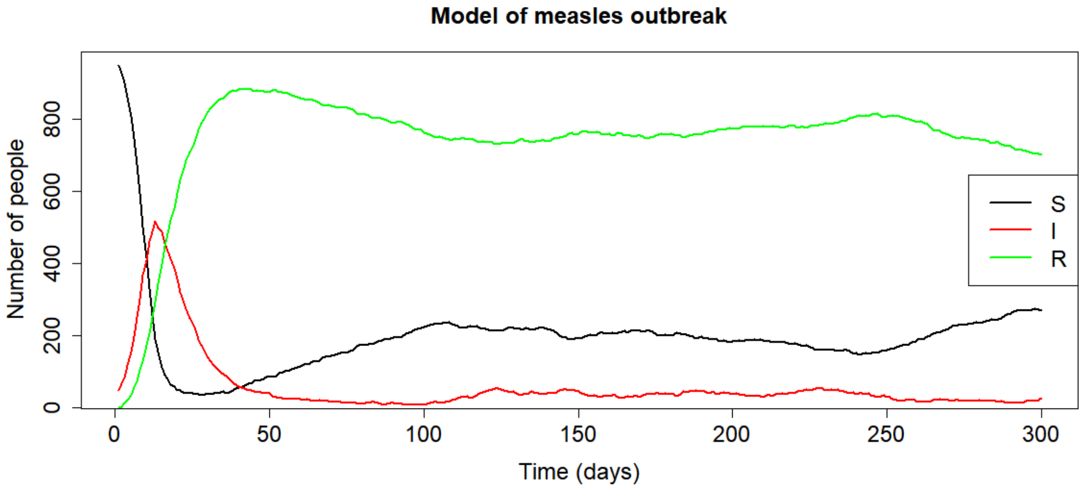


Figure 1: Temporal evolution of number of peoples in groups S, I, and R.

	Vaccinated individuals: 0	Vacc.: 100	Vacc.: 600	Vacc.: 800
Peak individuals	524	439	97	51
Time of Peak	12.85	13.69	16.86	2.05

Table 1: Expected maximum number of infected individuals and expected time at which the number of infected individuals first takes its highest value dependend on how many individuals are vaccinated.

g)

Our approach is the following: The number of vaccinated people are recovered or immune for the rest of their life, thus, they are in group R in the whole simulation. We take these people out of the simulation in the first step and simulate the time steps only for the remaining people, i.e., for 900 individuals, 400 individuals, and 200 individuals. In the end, we add the number of vaccinated people to the inviduals of the group R.

As more and more individuals get vaccinated, the outbreak of the disease is less severe. As the probability that a susceptible individual becomes infected depends on the number of already infected individuals, less severe outbreaks have a self-accelerating function. The pool of individuals who might be susceptible decreases and so does the number of infected individuals which then has a rebound effect. Via this channel, vaccinated individuals also protect unvaccinated individuals via a diminished spread of the disease. In the case of 600 and 800 vaccinated individuals in the population, even the extinction of the disease is achieved since if there are no infected individuals left at one time point, this observation will remain since no individual can catch the disease.

One can see that as more people become vaccinated, the maximum number of infected

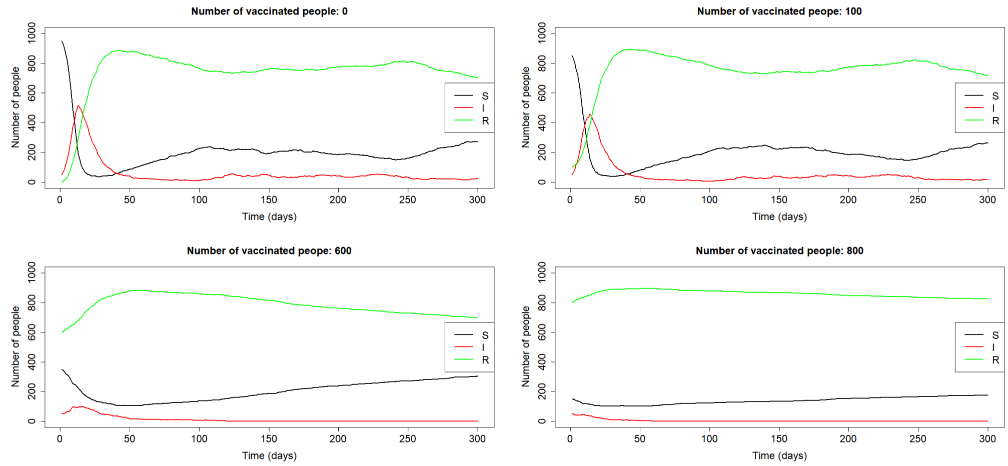


Figure 2: Temporal evolution of number of peoples in groups S, I, and R dependend on how many people were vaccinated initially.

individuals decreases in each case whereas this is not the case for the time point at which this happens. If 800 people are vaccinated, the effect of a constant probability that a person moves from the infectioned group moves to the recovered group dominates the relative probabiliy from the susceptible group nearly from the beginning. If no people are vaccinated, the peak number of individuals infected at one point is severe and sharp as people recover relatively fast. If individuals have recovered, it is relatively unlikely that they become susceptible fast. This is the reason why the outbreak distributed over a longer time window if some people are vaccinated: The pool of susceptible individuals is not pulled empty from the beginning.

Problem 2: Insurance claims

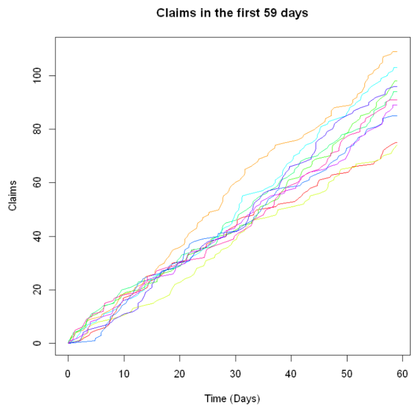
In this problem, we will use a Poisson process $X(t)$ that gives the number of claims in some time. Here, we will look at the period from January 1st to March 1st, which gives us exactly 59 days. For this process, we have that $\lambda = 1.5$.

a)

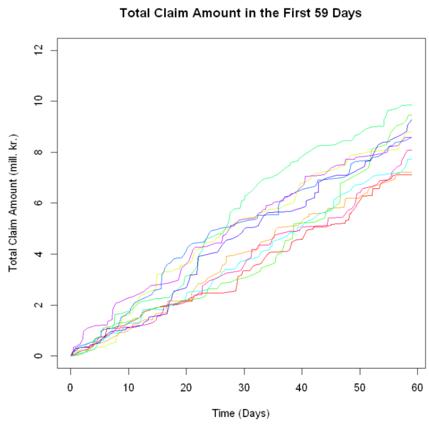
- Here, we are trying to determine $P(X(59) > 100)$, meaning that we want to find the probability of having more than 100 claims in the first 59 days.

We have that, for a Poisson process: $X(t) = X(t) - X(0) \sim \text{Poisson}(\lambda t)$. And,

$$P(X(59) > 100) = 1 - P(X(59) - X(0)) = 1 - \sum_{x=0}^{100} \frac{e^{-\lambda}\lambda^x}{x!} = 0.1028$$



(a) Simulations of $X(t)$ for $0 \leq t \leq 59$



(b) Simulations of $Z(t)$ for $0 \leq t \leq 59$

- After coding 1000 simulations we get that the probability is 0.108, giving us a very similar result to what we have found previously in this task.

Figure 3a shows a plot of 10 realizations of the simulation. From this we can clearly conclude that the most of the realizations can be found below 100 claims.

b)

The amount of money for each claim is denoted by C_i, which follows the exponential distribution with rate parameter $\gamma = 10$. We assume that the C_i are independent and independent of the claim arrival time. The total claim amount is $Z(t) = \sum_{i=1}^{X(t)} C_i$.

From the coding we find that the estimated probability of exceeding 8 million kr is 0.719, which is a fairly high number. However, one could already predict high probability, as the expected value of $Z(t)$ is 8.85. See the calculations for expectation below.

$$E[Z(t)] = E\left[E\left[\sum_{i=1}^{X(t)} C_i\right]\right] = E\left[\sum_{i=1}^{X(t)} \frac{1}{\gamma}\right] = \frac{1}{\gamma}E[X(t)] = \frac{1}{\gamma}\lambda t = 8.85$$

Figure 3b shows a plot of 10 simulations of $Z(t)$ in the first 59 days. We conclude that the amount of monetary claims rarely exceeds 10 million kr.

c)

To show that $Y(t)$ is a Poisson distribution we need to check if the function satisfies all the properties of a Poisson process.

1. $Y(0) = 0$: This is true since there are no claims to be analyzed when time is at 0.

2. $Y(t)$ has independent increments: In this case, the claim amounts are independent and also independent of the claim arrival times, meaning that the number of claims in any time period is also independent.

3. $Y(t)$ has stationary increments: The rate of the Poisson process $\lambda(t)$ is constant at 1.5, meaning that the distribution of the claim amounts only depends on the length of the interval, in any time frame.

4. $Y(t)$ has Poisson-distributed increments: We have that $\lambda(t)$ is 1.5, so the number of claims in any time interval $[s, t]$ follows a Poisson distribution with mean $1.5 \cdot (t - s)$.

From this, we can conclude that $Y(t)$ is a Poisson process.

Now we have to find the rate of $Y(t)$. For this, we need the mean number of claims analyzed per unit time. We have that λ is 1.5 from $X(t)$, so the mean number of claims per unit time is the same value, in this case.

However, we are only interested in claims that exceed 250,000 kr, so the rate of $Y(t)$ is equal to $\lambda_{X(t)}$ multiplied by the probability of a claim being higher than the value given. So:

$$\lambda_{Y(t)} = \lambda_{X(t)} \cdot P(C > 0.25)$$

Now we have to find $P(C > 0.25)$:

$$P(C > 0.25) = \int_{0.25}^{\infty} \gamma \cdot \exp(-\gamma \cdot x)\, dx = \int_{0.25}^{\infty} 10 \cdot \exp(-10 \cdot x)\, dx$$

Which gives us:

$$P(C > 0.25) = [-\exp(-10 \cdot x)]_{0.25}^{\infty} = \exp(-10 \cdot 0.25)$$

And we can then conclude that:

$$\lambda_{Y(t)} = 1.5 \cdot \exp(-\frac{5}{2})$$