Project 1 : Amandus Omholt Nygaard, Yawar Mahmood

Problem 1

a)

We then have three states in our state space $\Omega = \{0, 1, 2\}$. Then, it is given that, $P_{01} = \beta$, $P_{12} = \gamma$, and $P_{20} = \alpha$. Our transition probabilities should satisfy $P_{ij} \geq 0$ and $\sum_{\Omega} P_{ij} = 1, \forall i \in \Omega$. As it is not specified whether one can go from state 0 to 2, state 1 to 0 and state 2 to 1, it is reasonable to assume that these probabilities are equal to zero. With this in mind, and the conditions presented above, the probability transition matrix will be given by

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

We claim that X_n is a Markov process, that is, the output of the map function from the previous time-step is fed back as the input of the current time-step:

$P(X_n \in A_n | X_1 \in A_1, \dots X_{n-1} \in A_{n-1}) = P(X_n \in A_n | X_{n-1} \in A_{n-1})$.

This is a reasonable assumption as the further development of the disease will only depend on how many that are in the various states of our state space in the current time step.

b)

Using notation of Pinsky and Karlin, in our case, $k = 2$ gives

$$P^2 = \begin{bmatrix} + & + & 0 \\ 0 & + & + \\ + & 0 & + \end{bmatrix} \begin{bmatrix} + & + & 0 \\ 0 & + & + \\ + & 0 & + \end{bmatrix} = \begin{bmatrix} + & + & + \\ + & + & + \\ + & + & + \end{bmatrix},$$

which means that P is regular and has a limiting distribution. Found the limiting probabilities for each state in Ω by solving the following equations

$$\begin{aligned} \pi_0 &= \pi_0 P_{00} + \pi_1 P_{10} + \pi_2 P_{20} \\ \pi_1 &= \pi_0 P_{01} + \pi_1 P_{11} + \pi_2 P_{21} \\ \pi_2 &= \pi_0 P_{02} + \pi_1 P_{12} + \pi_2 P_{22} \\ \pi_0 + \pi_1 + \pi_2 &= 1 \end{aligned}$$

.

Discarding the equation for π_2 gives the system:

$$\begin{bmatrix} P_{00} - 1 - P_{20} & P_{10} - P_{20} \\ P_{01} - P_{21} & P_{11} - P_{21} - 1 \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix} = \begin{bmatrix} -P_{20} \\ -P_{21} \end{bmatrix}$$

$$\begin{bmatrix} -\beta - \alpha & -\alpha \\ \beta & -\gamma \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix} = \begin{bmatrix} -\alpha \\ 0 \end{bmatrix}$$

Let $\beta = 0.01$, $\gamma = 0.10$, and $\alpha = 0.005$. This yields the solution:

$$\pi = (\pi_0, \pi_1, \pi_2) = (\pi_0, \pi_1, 1 - \pi_0 - \pi_1) = \left(\frac{10}{31}, \frac{1}{31}, \frac{20}{31}\right)$$

Mean days in each state per year is then respectively given by $\frac{3650}{31}$, $\frac{365}{31}$, and $\frac{7300}{31}$.

c)

Let the individual be susceptible at time 0 ($X_0 = 0$). We simulate 20 years, 7300 time steps of the Markov chain X_n, using α, β, and γ as specified in b). The simulation is run in R with fixed seed equal to 1. This seed is used throughout this task. We run the realisation once.

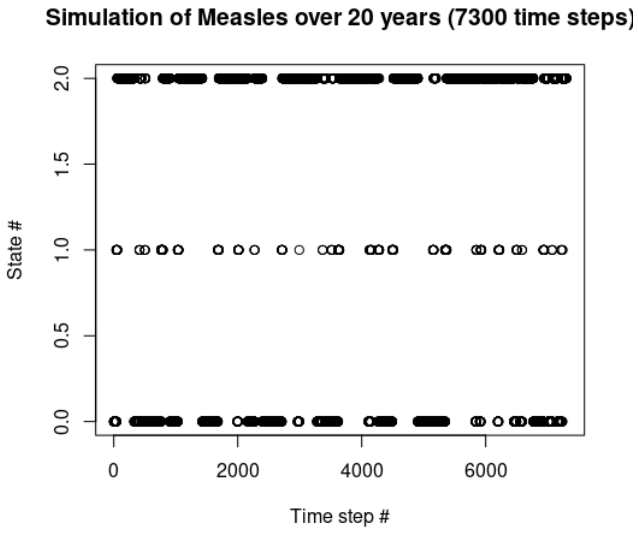

Figure 1: 1 realizations of Measles outbreak

For the realisation in Figure 1, we use the last 10 years of the 20 to estimate the limiting distribution. We ran the realisation 30 times. The limiting distributions are given in Figure 2. The blue lines are the upper and lower bounds for the CI, given in Table 1

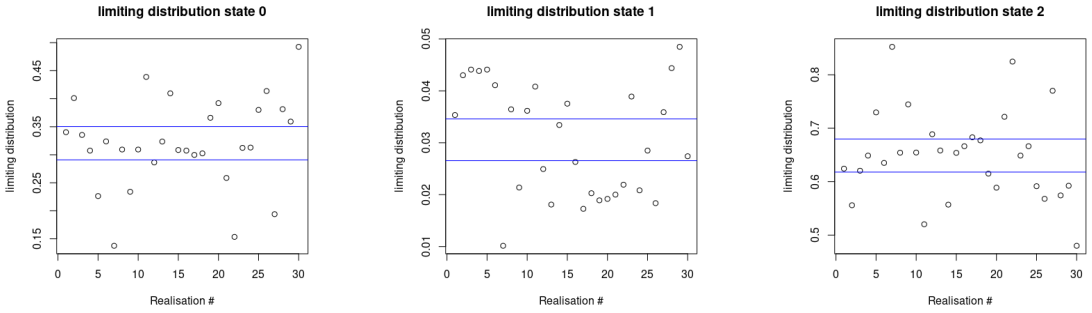

Figure 2: Limiting distribution for 30 realisations

The CIs was calculated by using:

$$CI = \hat{\pi} \pm t^* \frac{s}{\sqrt{n}}$$

Where $\hat{\pi}$ is the sample mean, s is the sample standard deviation, n is the sample size, and t^* is

Bound	π_0	π_1	π_2
Upper bound	0.3502854	0.03459189	0.6798301
Lower Bound	0.290847	0.02654054	0.6179051

Table 1: 95% CI for limiting distribution

the critical value of the student t-distribution with parameter n-1. Under the calculation, the assumption that the sample had a normal distribution was made.

We use the student t-distribution over a normal distribution. The reason behind this is that the t-distribution is used as an alternative to the normal distribution when sample sizes are small in order to estimate confidence.

From Figure2, we see that many of the values for π_0 and π_2 fall inside, or lay close to the CI. This, and the fact that the theoretical values for the limiting distributions fall inside the CI, indicate that the calculated values for the limiting distribution are correct. We further see that the values for π_1 for the most part fall outside the confidence interval. If we run the same test, and set our seed to 2, many more limiting distributions π_1 fall in the CI. It seems reasonable to conclude that this result is due to the seed.

d)

Let S_n be susceptible individuals, I_n be infected individuals and R_n be recovered individuals. Let $Z_n = (S_n, I_n)$ and $Y_n = (S_n, I_n, R_n)$.
Y_n is a Markov chain, since the state space is complete. That is, one can reach all states in our state space using the information provided in $S_n, I_n R_n$. The same is true for Z_n. By using S_n, I_n and the size of the population N, one can reach all of R_n, since $R_n = N - S_n - I_n$. Y_n and Z_n has the same state space. When it comes to I_n, one has too little information. One cannot find out how S_n and R_n will be in further states, and I_n is therefore not a Markov chain. The state space is also reduced.

e)

We want to simulate how the number of infected, recovered and susceptible, develop over a period of 300 days. We set the seed to be 1, and make a realisation given in Figure 3.

The Markov chain behaves differently in the interval 0-50 and 50-300. The reason for this is among other things the initial condition. The initial condition is set to $(S_0, I_0, R_0) \to (950, 50, 0)$. This is a rather extreme situation to start in, and is far from an equivalence state. Since the probability from being susceptible to being infected has the form $\beta = \frac{0.5 I_n}{N}$, the probability of becoming infected will increase rapidly with the number of infected. Since there are many susceptible to start with, almost the whole population size, the number of infected will grow fast in the start. But when the number of infected increases, it also becomes more likely that more will recover. This causes the big jump at the start to die out.

f)

Setting seed equal to one, and running 1000 simulations of the outbreak.

The severity of an outbreak can be interpreted through how quickly the number of infected people increase. If they increase rapidly over a short period of time, we say that the outbreak has high severity.

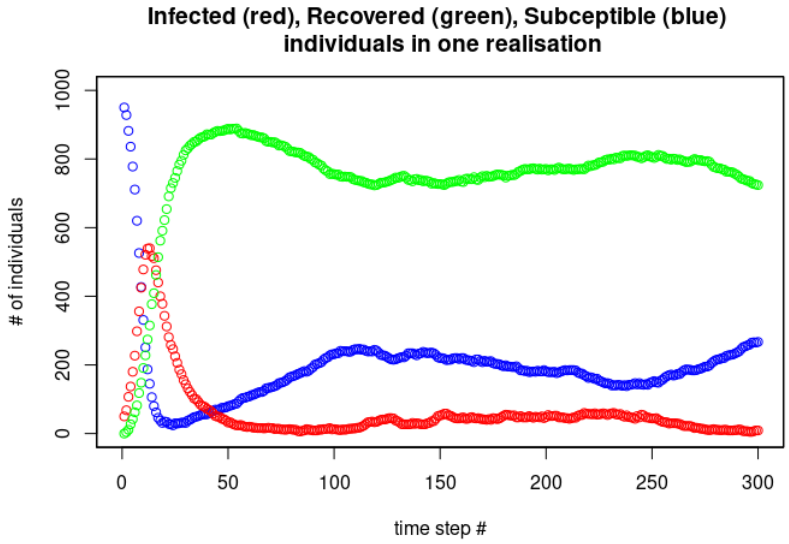

Figure 3: number of infected, recovered and susceptible in 300 days

Bound	$E[\max\{I_0, I_1, \ldots, _n\}]$	$E[\min\{\text{argmax}\{I_n\}\}]$
Upper bound	523.8307	521.2813
Lower Bound	12.95239	12.85161

Table 2: 95% CI $E[\max\{I_0, I_1, \ldots, _n\}]$ and $E[\min\{\text{argmax}\{I_n\}\}]$

We can say with 95 % confidence that the expected maximum number of infected individuals will lie in the interval given in Table 3. What is worrying is that the expected time at which the number of infected individuals first takes its highest value occurs 12.8 to 12.9 days after the outbreak. We will go from 50 infected to around 500 infected in a period of 12 days. That is an increase of nearly 900 % in a period of 12 days. The outbreak is therefore to be considered of high severity.

g)

Immune	Bound	$E[\max\{I_0, I_1, \ldots, _n\}]$	$E[\min\{\text{argmax}\{I_n\}\}]$
100	Upper bound	439.4947	437.1293
	Lower Bound	13.70908	13.58492
600	Upper bound	98.39431	96.82369
	Lower Bound	17.17922	16.62878
800	Upper bound	51.42904	51.16896
	Lower Bound	2.184934	1.959066

Table 3: 95% CI $E[\max\{I_0, I_1, \ldots, _n\}]$ and $E[\min\{\text{argmax}\{I_n\}\}]$ for various amount of immune individuals.

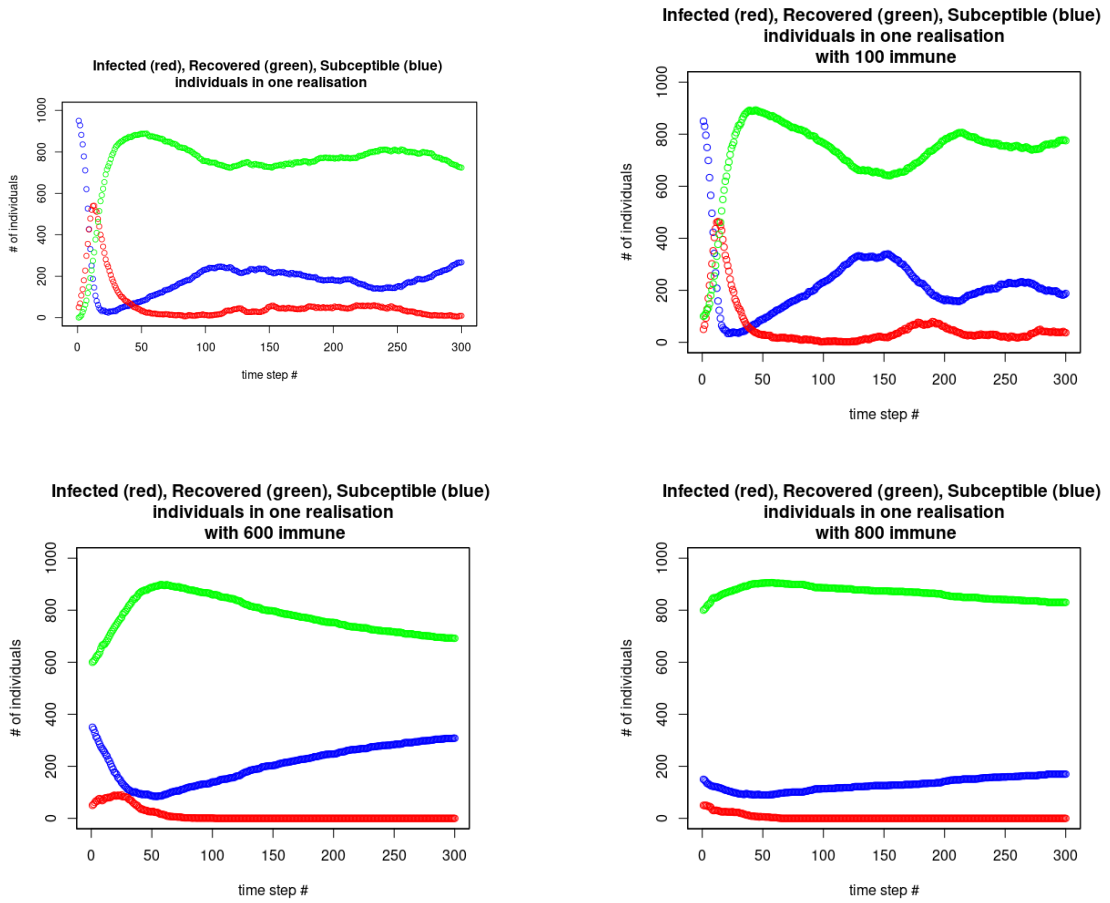

Figure 4: number of infected, recovered and susceptible in 300 days, with different immunity rates

As evident from Table 3, the expected maximum number of infected individuals fall with the number of immune in the population. The immune part of the population is part of the recovered individuals, but they are not affected by the disease. If we look at the probability of becoming infected, $\beta = \frac{0.5I_n}{N}$, we keep N constant, as it was before, but reduce I_n, in that there are fewer people who are susceptible. We reduce the value to β, fewer people get infected, and the spike becomes smaller, and dies out quicker.

Expected time at which the number of infected individuals first takes its highest value also increases. This is called flattening the curve, as you don't want to stress the hospitals with all the infected cases at once. When the immunity starts to approach the size of the population, we see that the expected time at which the number of infected individuals first takes its highest value approaches zero. This is evident in Table 3. The pandemic is over before it has even started.

Problem 2

a)

The Probability that there are more then 100 claims at March 1st at 00:00.00 ($t = 59$) is

$$P(X(59) > 100) = 1 - P(X(59) \leq 100) = 1 - \sum_{x=1}^{100} \frac{(\lambda t)^x}{x!} e^{-\lambda t} = 1 - \sum_{x=1}^{100} \frac{(1.5 \cdot 59)^x}{x!} e^{-1.5 \cdot 59}$$
$$= 0.1028.$$

To verify the result a function that simulates $X(t)$. This function was used to simulate 1000 realization for $X(t)$ on the interval $0 \leq t \leq 59$. The estimate was found by taking the number of simulation that ended with over 100 claims divided by total number of simulations, the result was $\hat{P}(X(t) > 100) = 0.103$, which is really close to the calculated value. 10 realizations are shown in Figure 5.

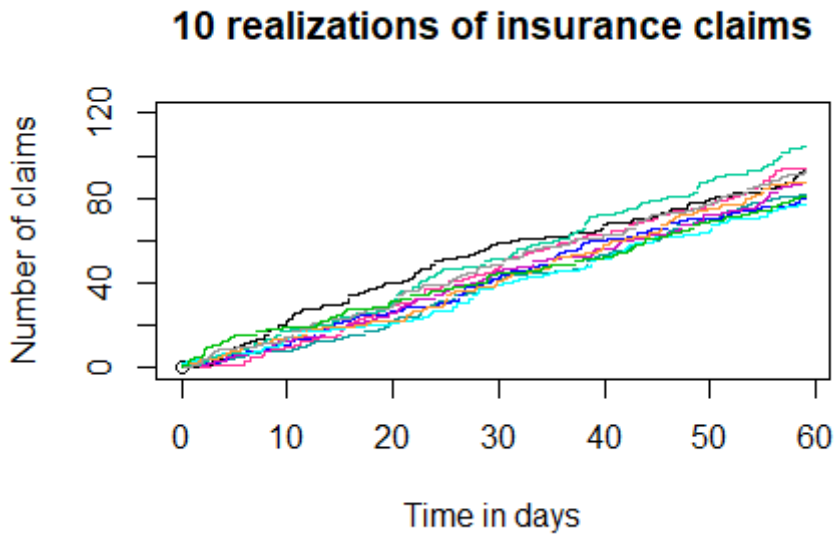

Figure 5: 10 realizations of number of claims

b)

A function that simulates the total claim amount was made by first simulating $X(t)$ with the function used in $2a$), and then simulate the claim amount for each sojourn time and add the amount from the previous time frames.
The claim amount $Z(t)$ for $0 \leq t \leq 59$ was simulated 1000 times and the frequency of the instances where the claim amount was above 8 million kr was used to estimate the probability, which gave $\hat{P}(Z(t) > 8) = 0.731$. 10 realizations are shown in figure 6.

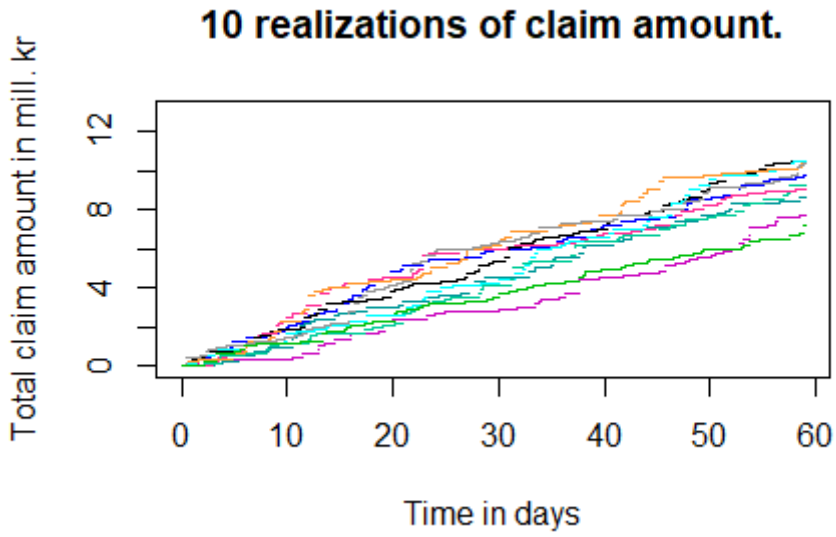

Figure 6: 10 realizations for claim amount

c)

Want to show that $Y(t)$ satisfy the following: 1) $Y(t)$ has independent increments. 2) $Y(s+t)-Y(s) \sim \text{Poisson}(\lambda_y t)$, for $s \geq 0, t > 0$ with some rate λ_y. 3) $Y(0) = 0$.
The third must be true as no insurance claim can come inn at $t = 0$, and thus no insurance claims need be investigated at $t = 0$.
$\{Y(t) : t \geq 0\}$ must have independent increments, since that amount of claims has independent increments and the claim amounts are independent and independent of arrival time. Let T_k be whether the claim need be investigated or not, which means that $T_k \sim \text{Bernoulli}(p_k)$. In this case $p_k = P(C_i > 0.25) = 1 - P(C_i \leq 0.25) = 0.0821$, since C_i has an exponential distribution with rate $\gamma = 10$. Thus, $\sum_{i=1}^{y} T_k \sim \text{Binomial}(y, p_k)$.

$$P(Y(s+t) - Y(s) = y) = P\left(\sum_{l=y}^{\infty}\left((X(s+t) - X(s) = l) \cap \sum_{i=1}^{l} T_k = y\right)\right)$$

$$= \sum_{l=y}^{\infty} P\left(X(s+t) - X(s) = l \cap \sum_{i=1}^{l} T_k = y\right) = \sum_{l=y}^{\infty} \frac{(\lambda t)^l}{l!} e^{-\lambda t} \binom{l}{y} p_k^y (1-p_k)^{l-y}$$

$$= \sum_{l=y}^{\infty} \frac{(\lambda t)^l}{l!} e^{-\lambda t} \frac{l!}{y!(l-y)!} p_k^y (1-p_k)^{l-y} = \frac{(\lambda p_k t)^y}{y!} e^{\lambda t} \sum_{l=y}^{\infty} \frac{(\lambda t)^{l-y}(1-p_k)^{l-y}}{(l-y)!}$$

$$= \frac{(\lambda p_k t)^y}{y!} e^{\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n (1-p_k)^n}{(n)!} = \frac{(\lambda p_k t)^y}{y!} e^{\lambda t} e^{\lambda t(1-p_k)} = \frac{(\lambda p_k t)^y}{y!} e^{\lambda_{p_k} t}$$

This means that $Y(s+t) - Y(s) \sim \text{Poisson}(\lambda_y t)$, with rate $\lambda_y = \lambda p_k = 1.5 \cdot 0.0821 = 0.123$.