June 3, 2024

Project 1: Aksel Stenvold, Casper Lindeman and Gonchigsuren Bor

Problem 1: Modelling an outbreak of measles

a)

In our universe, humans can have one of three states, susceptible (S), infected (I), and recovered and immune (R). Given the transition probability matrix

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

The P matrix in general describes the system

$$P = \begin{bmatrix} P(S|S) & P(I|S) & P(R|S) \\ P(S|I) & P(I|I) & P(R|I) \\ P(S|R) & P(I|R) & P(R|R) \end{bmatrix}$$

A visual representation is given in Figure 1. We can see that the sum of the possibilities from one state to every state is equal to one, which means it is a valid Markov chain.

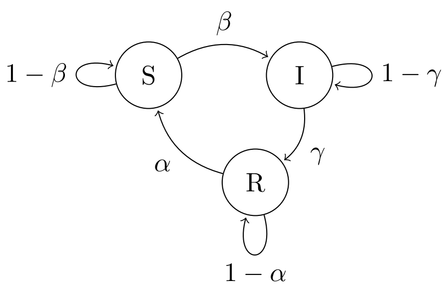

Figure 1: Markov Chain based on P

b)

We now assume that $\beta = 0.01, \gamma = 0.10$ and $\alpha = 0.005$. Which means that

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

For the Markov chain to have a limiting distribution, it has to be irreducible and aperiodic. Since our Markov chain is a closed circle, and $0 < \beta, \gamma, \alpha < 1$ it means that every possible state is reachable from any possible state given enough enumerations. Since our Markov chain is a closed circle, and $0 < \beta, \gamma, \alpha < 1$ it also means that the Markov chain is aperiodic.

Let's calculate the long-run mean number of days per year spent in each state.

$$P^T - I_3 = \begin{bmatrix} -\frac{1}{100} & 0 & \frac{1}{200} \\ \frac{1}{100} & -\frac{1}{10} & 0 \\ 0 & \frac{1}{10} & \frac{1}{200} \end{bmatrix} \sim \begin{bmatrix} -\frac{1}{100} & 0 & \frac{1}{200} \\ 0 & -\frac{1}{10} & \frac{1}{200} \\ 0 & 0 & 0 \end{bmatrix} \implies v = \begin{bmatrix} P(S) \\ P(I) \\ P(R) \end{bmatrix} = \begin{bmatrix} \frac{5}{16} \\ \frac{1}{16} \\ \frac{10}{16} \end{bmatrix}$$

If we assume 365 days a year the expected number of days in each state for an individual, rounded to the nearest whole day is

$$Days = \begin{bmatrix} \text{Days } (S) \\ \text{Days } (I) \\ \text{Days } (R) \end{bmatrix} = \begin{bmatrix} 114 \\ 23 \\ 228 \end{bmatrix}.$$

c)

We can write a code to simulate an individual over a period of 20 years, which is 7300 steps, and use the last 10 years to estimate the number of days spent in S, I, and R. The code runs 30 Markov Chains, then calculates the 2.5% and 97.5% of state proportions. These percentiles, multiplied by 365, give the 95% Confidence Intervals in days for each state. Our results says that $70.53 < S < 171.78$, $4.41 < I < 18.28$ and $183.19 < R < 288.56$. Based on our results we can see that the analytical value for S and R is somewhere within the interval, however, the actual value for I which is 23 does not fit inside the confidence interval we calculated using simulations. The deviation may be a result from too few samples.

d)

For a process to qualify as a Markov Chain, it must be memoryless. This means that the next state of the process should be dependent only on its current state and not on the sequence of states that preceded it.

Process for I:

Given:

$$I_{n+1} = S_n \beta_n + I_n(1-\gamma) \tag{1}$$

$$= S_n \cdot \frac{0.5 I_n}{N} + I_n(1-\gamma) \tag{2}$$

Conclusion: I_{n+1} is influenced by both S_n and I_n. Therefore, I is not a Markov Chain since its future state depends on more than just its immediate past.

Process for Z:

Given:

$$Z_{n+1} = (S_n(1-\beta_n) + R_n\alpha, S_n\beta_n + I_n(1-\gamma)) \tag{3}$$

Conclusion: Z_{n+1} is influenced by S_n, I_n, and R_n. Thus, Z is not a Markov Chain for the same reasons as above.

Process for Y:

Given:

$$Y_{n+1} = (S_n(1-\beta_n) + R_n\alpha, S_n\beta_n + I_n(1-\gamma), \gamma I_n + (1-\alpha)R_n) \tag{4}$$

$$= \left(S_n\left(1 - \frac{0.5I_n}{N}\right) + R_n\alpha, S_n\left(\frac{0.5I_n}{N}\right) + I_n(1-\gamma), \gamma I_n + (1-\alpha)R_n\right) \tag{5}$$

Where:

$$S_n = Y_n[1]$$
$$I_n = Y_n[2]$$
$$R_n = Y_n[3]$$

Conclusion: Y_{n+1} can be expressed entirely in terms of the components of Y_n. Therefore, Y qualifies as a Markov Chain since its next state is determined solely by its current state.

e)

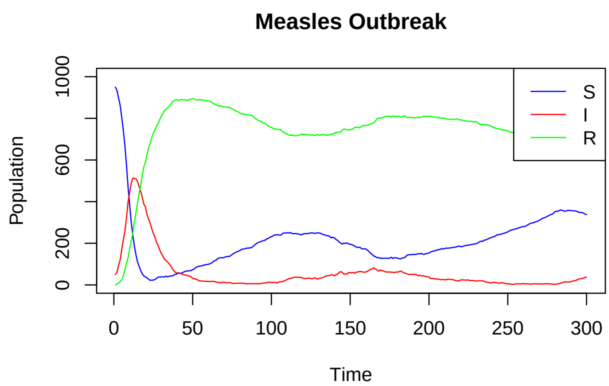

Figure 2: Exercise 1e: A simulation of a measles outbreak

In the initial phase, we observe a rapid increase in infections. This can be attributed to the large number of initially susceptible individuals combined with a significant number of infected individuals. There's a brief delay before the count of recovered individuals begins to climb. After 50 days, the Markov chain starts to stabilize, and we don't observe any drastic changes.

f)

The confidence interval for the maximum number of infected individuals suggests that, in 95% of the cases, we can expect between approximately 521.56 and 524.09 individuals to be infected at the peak of the outbreak. This gives us a relatively narrow range, indicating that the model predictions are quite consistent. The mean value of 522.826 individuals provides an estimate of the most likely scenario for the peak.

The outbreak is expected to reach this peak within approximately 12.876 days. The 95% confidence interval ranging from 12.82638 to 12.92562 days further solidifies this estimation. This early peaking suggests a rapid initial spread before it stabilizes.

g)

To simulate the impact of vaccinations on the population, we subtracted the number of vaccinated individuals from the susceptible group while maintaining the total population count constant.

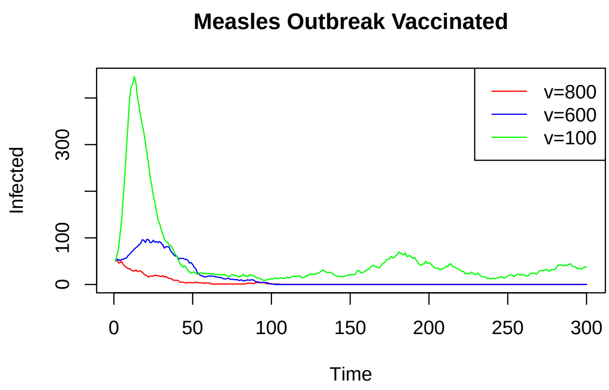

Figure 3: Exercise 1g: A simulation of a measles outbreak after vaccination.

From the graph, it's evident that as the number of vaccinated individuals rises, there's a significant reduction in the peak of infections and fewer overall infections.

In the study of peak infections ($I_{\max}$) under different vaccination scenarios, stark contrasts are evident:

- **800 Vaccinated:** $I_{\text{max1\_mean}} = 51.427$, 95% CI = [51.28184, 51.57216].
- **600 Vaccinated:** $I_{\text{max2\_mean}} = 97.077$, 95% CI = [96.28438, 97.86962].
- **100 Vaccinated:** $I_{\text{max3\_mean}} = 438.896$, 95% CI = [437.6395, 440.1525].

Comparing $I_{\max}$ across scenarios reveals a clear trend: higher vaccination rates correlate with a pronounced reduction in peak infections, underscoring the impact of vaccinations in mitigating outbreak severity.

Problem 2: Insurance claims

a)

Given that we have a Poisson process, we can compute the probability that there will be more than 100 after $t = 59$ days.

$$P(X > 100) = 1 - P(X100) \tag{6}$$
$$= 1 - \sum_{x=0}^{100} P(X = x) \tag{7}$$
$$= 1 - \sum_{x=0}^{100} \frac{(\lambda t)^x e^{-\lambda t}}{x!}, \tag{8}$$

and with $t = 59$ the equation above results in

$$P(X > 100) = 1 - \sum_{x=0}^{100} \frac{(-1.5 \cdot 59)^x e^{-1.5 \cdot 59}}{x!} = 0.10282, \tag{9}$$

meaning that the probability of getting more than 100 claims on March 1st, given that $X(t)$ is a Poisson process, is at around 10.282%.

To verify the calculations, we have written a code in R that simulates 1000 realisations of the Poisson process. This yields us a result that ranges between 0.8 and 1.2. Note that we are working with random variables, meaning that it is natural to have deviation from the theoretical value.

Now, let's plot a figure that shows 10 realisations of $X(t), 0 \leq t \leq 59$, plotted in the same figure.

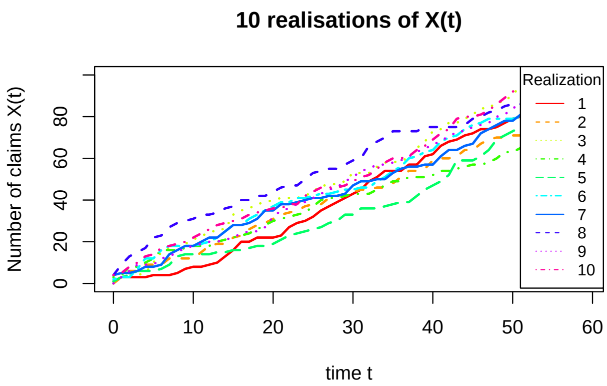

Figure 4: Exercise 2a: 10 realisations of $X(t)$

b)

Writing a code that makes 1000 simulations to estimate the probability that the total claim amount exceed 8 mill. kr. after 59 days, we get

$$P(Z > 8) \approx 0.748, \tag{10}$$

for the estimated probability for $Z > 8$. Note that, here again, the result varies from simulation to simulation as we are working with random variables that are randomly generated for every simulation.

Now, let's make a plot that shows 10 realisations of $Z(t), 0 \leq t \leq 59$ plotted in the same figure.

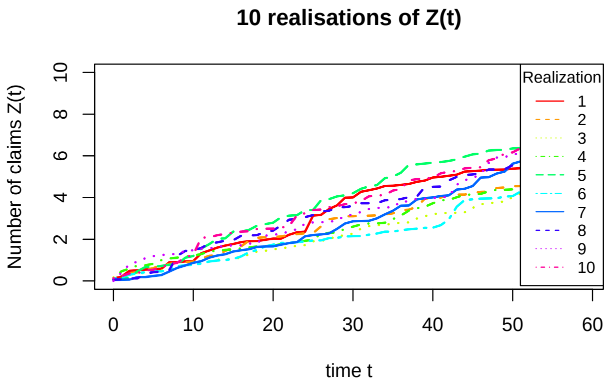

Figure 5: Exercise 2b: 10 realisations of $Z(t)$.

c)

Let Y_t denote the number of claims that exceed 250'000kr, and let Y_t' be the complement of Y_t, or in other words, let Y_t' denote the number of claims less than 250'000 kr. Since X_t denotes the number of total claims and is, this would mean that

$$X_t = Y_t + Y_t'. \tag{11}$$

Furthermore, we can recall that each claim is independent, as well as the probability of each claim happening does not vary no matter how many claims are made. This means that we can say that each claim that succeeds 250'000kr can be seen as a success, meaning that we have a Bernoulli trial and that Y_t can be modeled by a binomial process. Now, using the law of total probability, we get

$$
\begin{aligned}
P(Y_t = k) &= \sum_{n=0}^{\infty} P(Y_t = k, X_t = n) && (12)\\
&= \sum_{n=0}^{\infty} P(Y_t = k | X_t = n) P(X_t = n) = && (13)\\
&= \sum_{n=0}^{\infty} P(Y_t = k) P(X_t = n) = && (14)\\
&= \sum_{n=0}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{(\lambda t)^n e^{-\lambda t}}{n!} && (15)\\
&= \frac{e^{-\lambda t (\lambda t p)^k}}{k!} \sum_{n=k}^{\infty} \frac{\mu (1-p)^{n-k}}{(n-k)!}, && (16)
\end{aligned}
$$

and using the Taylor series for e^x on the sum part of the equation, we get

$$P(Y_t = k) = \frac{e^{-\lambda t(\lambda tp)^k}}{k!} e^{\lambda t(1-p)} \tag{17}$$

$$= \frac{e^{-\lambda pt}}{k!} (\lambda tp)^k, \tag{18}$$

which is the claimed Poisson distribution, meaning that $Y_t \sim \text{Pois}(\lambda tp)$.

Now, note that the value for p, that is the average probability for a claim to be over 250'000kr is, as of now, unknown. Since we know that the claims $C_i \sim \text{Exp}(\gamma)$, where $\gamma = 10$, p is therefore

$$p = 1 - P(C \leq 0.25) = 1 - \int_0^{0.25} 10e^{-10x} \mathrm{d}x \approx 0.082085, \tag{19}$$

meaning that the rate λp is equal to $\lambda p = 1.5 \cdot 0.082085 \approx 0.1231$.