TMA4265 Stochastic Modelling - Fall 2020

Project 1

Alexander J Ohrt, Jim Totland

September 2020

Problem 1: Modelling the outbreak of measles

a) $\{X_n : n = 0, 1, \ldots\}$ is a Markov chain because it satisfies the Markov property, i.e. the current state only depends on the previous state

$$\begin{aligned} &P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\ &\quad = P(X_{n+1} = j | X_n = i) \qquad n = 0, 1, \ldots \quad \forall i, j \in \{\mathrm{S}, \mathrm{I}, \mathrm{R}\}. \end{aligned} \tag{1}$$

We let the states S (susceptible), I (infected) and R (recovered and immune) correspond to states 0, 1 and 2, respectively. We have probability β to transition from 0 to 1, probability γ to transition from 1 to 2, and probability α to transition from 2 to 0. With no other allowed transitions between states, except the transition from a state to itself, the transition probability matrix necessarily becomes

$$\mathbf{P} = \begin{pmatrix} 1-\beta & \beta & 0 \\ 0 & 1-\gamma & \gamma \\ \alpha & 0 & 1-\alpha \end{pmatrix}. \tag{2}$$

b) The answers to the questions in the bullet points follow.

- This Markov chain is irreducible, because all the states communicate. More precisely, 1 is accessible from 0, 2 is accessible from 1 and 0 is accessible from 2, which means that all states communicate.
- The point above implies that there is only one equivalence class: $\{0, 1, 2\}$. All states are recurrent, since one cannot escape this equivalence class, which has a finite number of states. More rigorously, it can easily be shown that the Markov chain is regular by computing $\mathbf{P}^2$. Then, for any $i \in \{0, 1, 2\}$, we have

$$\lim_{n\to\infty} P_{ii}^{(n)} = \pi_i > 0. \tag{3}$$

  Hence, $\sum_{i=1}^{\infty} P_{ii}^{(n)} = \infty$, which implies that the states are recurrent.

- Since the diagonal elements in the transitional probability matrix are all strictly positive, it is always possible to transition to the current state in the next step, which implies that for any state $i \in \{0, 1, 2\}$ the periodicity is

$$d(i) = \gcd\{n \geq 1 : P_{ii}^{(n)} > 0\} = \gcd\{1, 2, 3, \ldots\} = 1, \tag{4}$$

which means that the Markov chain is aperiodic. It would suffice to show this for one state, as the periodicity is a property of the equivalence class.

c) In this task we define three stochastic variables

$$\begin{aligned}
T_1 &= \min\{n \geq 1 : X_n = 1\}, \\
T_2 &= \min\{n \geq 1 : X_n = 2, X_{n-1} = 1\}, \\
T_3 &= \min\{n \geq 1 : X_n = 0, X_{n-1} = 2\},
\end{aligned}$$

where $n = 1, 2, \ldots$ are the steps taken forward in time.
The expected values are given by

- $\mathrm{E}[T_1 | X_0 = 0] = \frac{1}{\beta} = 20$,
- $\mathrm{E}[T_2 | X_0 = 0] = \frac{1}{\gamma} = 30$,
- $\mathrm{E}[T_3 | X_0 = 0] = \frac{1}{\beta} + \frac{1}{\gamma} + \frac{1}{\alpha} = 130$,

where the numerical values are consequences of the given probabilities $\beta = 0.05$, $\gamma = 0.10$ and $\alpha = 0.01$.

The justification of the answers is in the following, utilizing first step analysis. The first bullet point is given by

$$v = (1 - \beta)v + \beta \cdot 0 + 1 \implies v = \frac{1}{\beta} = 20,$$

where $v = \mathrm{E}[T_1 | X_0 = 0]$.

Next, we define the stochastic variable $v_i = E[T_2 | X_0 = i]$ for $i = 0, 1$. The expected value in question is therefore calculated from the system

$$\begin{aligned}
v_0 &= (1 - \beta)v_0 + \beta v_1 + 1 \\
v_1 &= (1 - \gamma)v_1 + \gamma \cdot 0 + 1,
\end{aligned}$$

which gives $v_0 = \frac{1}{\beta} + \frac{1}{\gamma} = 30$.

Lastly, we define the stochastic variable $v_i = E[T_3 | X_0 = i]$ for $i = 0, 1, 2$. The expected time to complete a full cycle through the states if the individual is in state 0 at time 0 is given by the system of equations

$$
\begin{aligned}
v_0 &= (1-\beta)v_0 + \beta v_1 + 1 \\
v_1 &= (1-\gamma)v_1 + \gamma v_2 + 1 \\
v_2 &= \alpha \cdot 0 + (1-\alpha)v_2 + 1,
\end{aligned}
$$

which gives $v_0 = \frac{1}{\beta} + \frac{1}{\gamma} + \frac{1}{\alpha} = 130$.

d) The results from the simulations are good estimations of the theoretical values. For example, one realization of taking the mean over 100 simulations of 50 years (18250 time steps) of the Markov chain gives the results 20.107, 30.099 and 129.88.

e) $\{I_n : n = 0, 1, \ldots\}$ is not a Markov chain. To show this explicitly, let $I_+^{(n)} \sim \mathrm{Bin}(n = S_{n-1}, p = \beta)$ denote the number of new infected individuals in state n. We similarly define $S_+^{(n)}$ and $R_+^{(n)}$, according to the transition probabilities. Hence,

$$
\begin{aligned}
&P(I_n = j | I_{n-1} = i, \ldots, I_0 = i_0) \\
&\quad = P(I_{n-1} + I_+^{(n)} - R_+^{(n)} = j) \\
&\quad \neq P(I_n = j | I_{n-1} = i),
\end{aligned} \tag{5}
$$

because $I_+^{(n)}$ has parameter $n = S_{n-1}$ (not to be confused with state number n), which is not known from the previous state in the Markov chain. This means that $\{I_n : n = 0, 1, \ldots\}$ is not a Markov chain. On the contrary, $\{Z_n = (S_n, I_n) : n = 0, 1, \ldots\}$ is a Markov chain, because

$$
\begin{aligned}
&P((S_n, I_n) = (i, j) | (S_{n-1}, I_{n-1}) = (i_{n-1}, j_{n-1}), \ldots, (S_0, I_0) = (i_0, j_0)) \\
&\quad = P(S_{n-1} + S_+^{(n)} - I_+^{(n)} = i, I_{n-1} + I_+^{(n)} - R_+^{(n)} = j) \\
&\quad = P((S_n, I_n) = (i, j) | (S_{n-1}, I_{n-1}) = (i_{n-1}, j_{n-1})).
\end{aligned} \tag{6}
$$

The last equality follows from the fact that all parameters involved in the stochastic variables in the second line are known. For example, $S_+^{(n)}$ has parameter $n = R_{n-1} = N - S_{n-1} - I_{n-1}$, where N is the total population.

f) One realization can be seen in figure 1. The reason behind the difference in behavior of the Markov chain in intervals $0 \leq n \leq 50$ and $50 < n \leq 300$ is that the Markov chain approaches its limiting distribution rapidly in the first time interval, while after this point, the distribution of individuals more or less only fluctuate around the limiting distribution.

It can be shown, by transition probability matrix multiplication, that this Markov chain is regular, which shows that the limiting probability exists. In fact, all values of $\mathbf{P}^2$ are strictly greater than 0.

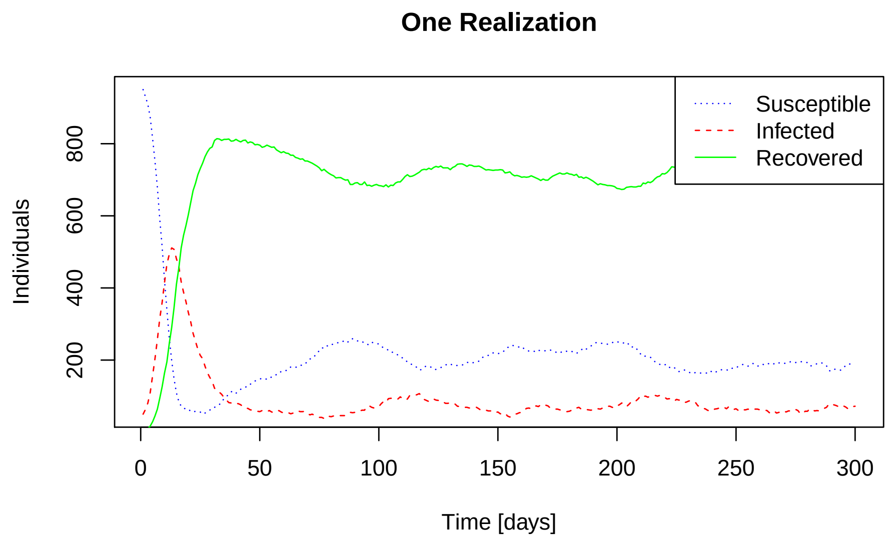

Figure 1: One realization of the Markov chain simulated until time step $n = 300$. The temporal evolutions of S_n, I_n and R_n are shown in the figure.

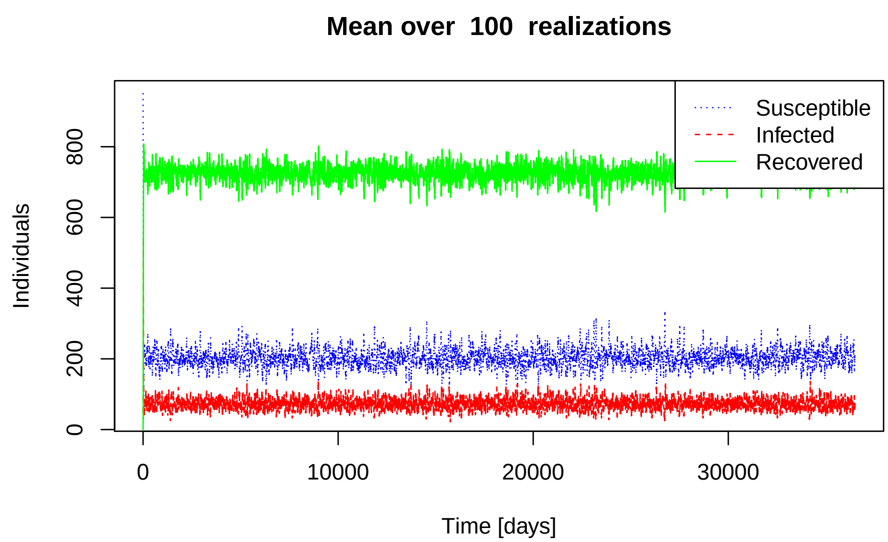

Figure 2: Mean over 100 realizations of the Markov chain $\{Y_n : n = 0, 1, 2 \dots\}$ simulated 36500 days into the future.

g) The mean over 100 realizations of the Markov chain simulated for 100 years is shown in figure 2. The approximate proportion of people in each state, calculated from the mean of the realizations, is $\bar{S} = 20.6\%$, $\bar{I} = 7.2\%$, and $\bar{R} = 72.2\%$, where $\bar{S}$, $\bar{I}$, $\bar{R}$ denote the proportion of susceptible, infected, and recovered, respectively.

h) $\mathrm{E}_1 = \mathrm{E}[\max\{I_0, \ldots, I_{300}\}] \approx 525$ and $\mathrm{E}_2 = \mathrm{E}[\min\{\arg\max_{n\leq 300}\{I_n\}\}] \approx 13$ are the approximated values in R. It is reasonable to think that the potential severity of the epidemic grows proportionally with E_1, since a higher value means more people will be sick at the same time, which leads to more strain on the health care system. Furthermore, the smaller the value of E_2, the less time the society has to prepare for the peak, which could potentially cause the outbreak to be more severe.

Problem 2: Insurance claims

a) If we let $\tilde{X} = X(59) - X(0) \sim \text{Poisson}(59 \cdot \lambda)$, the probability that there are more than 100 claims in 59 days is given by

$$P(\tilde{X} > 100) = 1 - P(\tilde{X} \leq 100) = 0.103. \tag{7}$$

Equivalently, we can use the waiting times of the events. Let $W_n \sim \text{Gamma}(n, \lambda)$ denote the time until the n^{th} jump. Then the desired probability is

$$\begin{aligned} W_{101} &\sim \text{Gamma}(\alpha = n = 101, \lambda = 1.5) \\ P(W_{101} \leq 59) &= \int_0^{59} \frac{1.5^{101}}{\Gamma(101)} x^{101-1} \exp(-1.5x) dx \approx 0.103. \end{aligned} \tag{8}$$

These calculations are verified by the simulated values, based on 1000 realizations from the Poisson process, $P(\tilde{X} > 100) \approx 0.102$. Figure 3 shows 10 realizations of $X(t), 0 \leq t \leq 59$.

b) The expected total claim amount is given by

$$\begin{aligned} \mathrm{E}[Z(59)] &= \mathrm{E}\left[\sum_{i=1}^{\tilde{X}} C_i\right] = \mathrm{E}\left[\mathrm{E}\left[\sum_{i=1}^{\tilde{X}} C_i | \tilde{X}\right]\right] \\ &= \mathrm{E}\left[\sum_{i=1}^{\tilde{X}} \mathrm{E}[C_i | \tilde{X}]\right] = \mathrm{E}\left[\sum_{i=1}^{\tilde{X}} \frac{1}{10}\right] \\ &= \mathrm{E}\left[\tilde{X}\frac{1}{10}\right] = \frac{1}{10}\mathrm{E}[\tilde{X}] = \frac{1.5 \cdot 59}{10} = 8.85. \end{aligned} \tag{9}$$

The variance of the total claim amount is given by

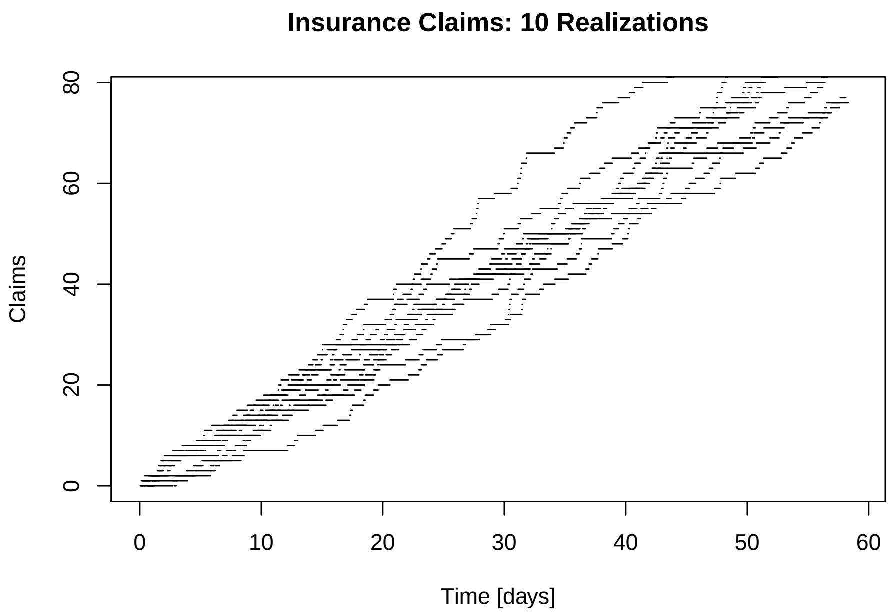

Figure 3: 10 realizations of the Poisson process for $0 \le t \le 59$.

$$\begin{aligned}
\mathrm{Var}[Z(59)] &= \mathrm{Var}\left[\sum_{i=1}^{\tilde{X}} C_i\right] \\
&= \mathrm{E}\left[\mathrm{Var}\left[\sum_{i=1}^{\tilde{X}} C_i|\tilde{X}\right]\right] + \mathrm{Var}\left[\mathrm{E}\left[\sum_{i=1}^{\tilde{X}} C_i|\tilde{X}\right]\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{\tilde{X}} \mathrm{Var}[C_i|\tilde{X}]\right] + \mathrm{Var}\left[\sum_{i=1}^{\tilde{X}} \mathrm{E}[C_i|\tilde{X}]\right] \\
&= \mathrm{E}\left[\frac{\tilde{X}}{10^2}\right] + \mathrm{Var}\left[\frac{\tilde{X}}{10}\right] = 2 \cdot \frac{1.5 \cdot 59}{10^2} \approx 1.77, \qquad (10)
\end{aligned}$$

where the third equality holds because of independent claims, which means that the covariance between the claims is zero. Estimations based on 1000 computer simulations give $\mathrm{E}[Z(59)] \approx 8.85$ and $\mathrm{Var}[Z(59)] \approx 1.76$, which are comparable to the theoretical values.