TMA4265 Stochastic Modelling – Fall 2021

Project 1

Background information

- The deadline for the project is Tuesday October 12 at 23:59.
- This project counts 10% of the final mark in the course.
- This project must be passed to be admitted to the final exam, and a reasonable attempt must be made for each problem to pass this project.
- The project should be done in groups of **two** or **three** people. You must sign up as a group in Blackboard before submitting your report and code.
- The project report should preferably be prepared using LaTeX and **must**:
  - be a pdf-file.
  - include necessary equations and explanation to justify the answers in each problem.
  - include the required plots in the pdf and reference them in the text.

  You do not need to repeat the question text in the report you submit.
- The computer code should be submitted as a separate file, and **not** as an appendix in the report. If you include any code in the report it should be as small excerpts which serve as natural parts of your answer to the question. Make sure your code runs. We may test it.
- There is a **7 page limit** for the project report. If you submit a longer report, we may not read it. The computer code is submitted as a separate file and may be as long as necessary.
- Make your computer code readable and add comments that describe what the code is doing.
- You are free to use any programming language you want as long as the code is readable, but you can only expect to receive help with `R`, `MATLAB` and `Python`.
- We will provide help with the project in the exercise classes in weeks 40 and 41, and in the lectures in week 40.
- If you have questions outside the aforementioned times, please contact the teaching assistants: `susan.anyosa@ntnu.no`, `silius.m.vandeskog@ntnu.no` or `kwaku.p.adjei@ntnu.no`.
- The pdf-file with the report and the files with computer code must be submitted through our Blackboard pages under "Projects". You need to sign up as a group before you can submit your answer.

Problem 1: Modelling an outbreak of measles

Throughout this problem you may assume that one year has exactly 365 days and that we are considering a population consisting of a fixed number $N > 0$ individuals. We use a simplified model where each individual only has three possible states: susceptible (S), infected (I), and recovered and immune (R). We model on a daily scale and let $n = 0, 1, \ldots$ denote time measured in days. We treat measles as an infectious disease, and we assume that each day

1) a susceptible individual can become infected or remain susceptible,

2) an infected individual can become recovered or remain infected,

3) a recovered individual can lose immunity and become susceptible, or remain recovered.

In the first stage of modelling, we assume that the individuals in the population are independent, and assume that each day, any susceptible individual has a probability $0 < \beta < 1$ of becoming infected tomorrow, any infected individual has a probability $0 < \gamma < 1$ of becoming recovered tomorrow, and any recovered individual has a probability $0 < \alpha < 1$ of losing immunity tomorrow.

a) Consider one specific individual, and let X_n denote the state of that individual at time n. Let the states 0, 1 and 2 correspond to S, I and R, respectively, and assume that $X_0 = 0$. Justify that $\{X_n : n = 0, 1, \ldots\}$ is a Markov chain and explain why 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) Assume that $\beta = 0.01$, $\gamma = 0.10$ and $\alpha = 0.005$. Explain why you can be certain that this Markov chain has a limiting distribution, and calculate (by hand) the long-run mean number of days per year spent in each state.

c) Assume that the individual is susceptible at time 0, and complete the following tasks:

- (Computer code) Simulate the Markov chain for 20 years (7300 time steps).
- (Computer code) For each run of this code, use the last 10 years (of the 20 years) to get an estimate of each of the quantities in **b)**.
- (Computer Code) Run the code 30 times and compute an approximate 95% confidence interval (CI) for each of the quantities in **b)** (using the 30 independent estimates).
- (Report) Explain briefly how you computed the CIs, provide the computed CIs, and discuss whether the computed CIs are compatible with your exact calculations in **b)**?

Due to the highly infectious nature of measles, the proportions of susceptible, infected and recovered individuals in the population will change with time. This in turn means that it is highly unrealistic to assume that β does not change with time. Assume that the population size $N = 1000$ is constant through time and at each time step consists of S_n susceptible individuals, I_n infected individuals, and R_n recovered individuals.

Assume that for each time step n, the probability that a susceptible individual becomes infected is $\beta_n = 0.5 I_n/N$, the probability that an infected individual recovers is $\gamma = 0.10$, and the probability that a recovered individual becomes susceptible is $\alpha = 0.005$. Assume that the $N = 1000$ individuals change states independently of each other at each time step given the values of β_n, γ and α. Define the discrete-time stochastic process $\{Y_n : n = 0, 1, \ldots\}$, where $Y_n = (S_n, I_n, R_n)$.

d) Let $Z_n = (S_n, I_n)$ for $n = 0, 1, \ldots$. Are $\{I_n : n = 0, 1, \ldots\}$, $\{Z_n : n = 0, 1, \ldots\}$, and $\{Y_n : n = 0, 1, \ldots\}$ Markov chains? Give definite arguments for all three.

Introduce 50 infected individuals in the population at time $n = 0$ by $Y_0 = (950, 50, 0)$.

e) Complete the following tasks:

- (Computer code) Simulate $\{Y_n : n = 0, 1, 2, \ldots\}$ until time step $n = 300$.
  Hint: At each time step, the number of new susceptible individuals, new infected individuals, and new susceptible individuals are all binomial distributions. After you determine the number of trials and the success probability for each of them, you can simulate them, for example, using `rbinom` in R.

- (Report) Choose one realization and show the temporal evolutions of S_n, I_n and R_n together in one figure. Give a short discussion of why the behaviour of the Markov chain is very different in time intervals 0–50 and 50–300.

f) A major interest in the modelling of infectious diseases lies in the explosive behaviour during the initial outbreak of the disease. Based on 1000 simulations of the outbreak for time steps $n = 0, 1, \ldots, 300$:

- (Computer code) Estimate the expected maximum number of infected individuals during the simulated time steps, $\mathrm{E}[\max\{I_0, I_1, \ldots, I_{300}\}]$, and the expected time at which the number of infected individuals first takes its highest value, $\mathrm{E}[\min\{\arg \max_{n \leq 300} \{I_n\}\}]$.

- (Computer code) Compute approximate 95% CIs for the two expected values.

- (Report) Provide the computed CIs, and discuss how would you use them assess the potential severity of the outbreak?

g) This question is open to different solutions, but you need to justify and describe the approach you choose in your report.

A strategy to avoid large outbreaks is immunisation programmes. Assume vaccination provides life-long immunity, and consider three cases: 100 individuals are vaccinated, 600 individuals are vaccinated, and 800 individuals are vaccinated.

We want to introduce 50 infected individuals among the unvaccinated individuals, and study the behaviour of the number of infected individuals through time. Modify the above model to accommodate the vaccinated individuals, and use your modified model to generate four realizations of the temporal evolution of the number of infected individuals in the same figure: the three cases described above, and the original case of no vaccinated individuals. Discuss what changes as more and more individuals are vaccinated.

Compute and discuss changes in the estimated expected values from **1f)** in the three new cases compared to the original case of no vaccinated individuals.

Problem 2: Insurance claims

Let $X(t)$ denote the number of claims received by an insurance company in the time interval $[0, t]$. We will assume that $\{X(t) : t \geq 0\}$ can be modelled as a Poisson process, where t is measured in days since January 1st at 00:00.00. Assume that the rate of the Poisson process is given by $\lambda(t) = 1.5$, $t \geq 0$.

a) Complete the following tasks:

- (Report) Compute the probability that there are more than 100 claims at March 1 at 00:00.00 (59 days).
- (Computer code) Write code to verify the calculation by simulating 1000 realizations of the Poisson process.
- (Report) Compare the estimated probability with the exact calculation and comment. Further, make a figure that shows 10 realizations of $X(t)$, $0 \leq t \leq 59$, plotted in the same figure.

Assume that the monetary claims are independent, and are independent of the claim arrival times. Each claim amount (in mill. kr.) has an exponential distribution with rate parameter $\gamma = 10$. This means that claim $C_i \sim \text{Exp}(\gamma)$, $i = 1, 2, \ldots$. The total claim amount at time t is defined by $Z(t) = \sum_{i=1}^{X(t)} C_i$.

NB: *Since γ is a* ***rate*** *parameter, the exponential distribution is parametrized as $f(c) = \gamma e^{-\gamma c}$, $c > 0$. This may differ from what you have seen in other courses, but we will exclusively use this parametrization of the exponential distribution in this course.*

b) Complete the following tasks:

- (Computer code) Write code that uses 1000 simulations to estimate the probability that the total claim amount exceeds 8 mill. kr. at March 1 at 00:00.00 (59 days).
- (Report) Provide the estimated probability, and make a figure that shows 10 realizations of $Z(t)$, $0 \leq t \leq 59$, plotted in the same figure.

c) The policy of the insurance company is to investigate a claim if and only if it exceeds 250000 kr.. For $t \geq 0$, let Y_t denote the number of claims, which need to be investigated, received in the time interval $[0, t]$. Prove that $\{Y(t) : t \geq 0\}$ is a Poisson process and compute its rate.