TMA4265 Stochastic Modelling – Fall 2020
Project 1

Background information

- The deadline for the project is Sunday October 4 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.
- 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 be a pdf-file that includes necessary equations and explanation to justify the answer 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. Make sure this code runs. We may test it.
- There is a **6 page limit** for the project report. If you submit a longer report, we may not read it. The 6 page limit does not include the computer code, which should be submitted as a separate file.
- 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 physical guidance in R2 during the exercise classes in week 39 and 40. There are **no** lectures in week 40, but you can receive physical guidance during lecture hours on Monday and digital guidance during lecture hours on Wednesday. Information about this will be posted on Blackboard.
- If you have questions outside the aforementioned times, please contact the teaching assistants `susan.anyosa@ntnu.no` or `mina.spremic@ntnu.no`.
- The pdf-file with the report and the files with computer code should 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 the outbreak of measles

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 $X_0 = 0$. Explain why $\{X_n : n = 0, 1, \ldots\}$ is a Markov chain and explain why the transition probability matrix is

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

b) Answer the following questions about $\{X_n : n = 0, 1, \ldots\}$.

- Is this a reducible or irreducible Markov chain?
- Determine the equivalence classes and determine whether they are recurrent or transient.
- What is the period of each state?

c) Assume that $\beta = 0.05$, $\gamma = 0.10$ and $\alpha = 0.01$. Calculate (by hand) the following expected values.

- The expected time until a susceptible individual becomes infected.
- The expected time until a susceptible individual becomes recovered.
- The expected time to complete a full cycle through the states if the individual is susceptible at time 0. *(I.e., let $T = \min\{n \geq 1 : X_n = 0, X_{n-1} = 3\}$ and calculate $\mathrm{E}[T|X_0 = 0]$.)*

d) Assume that one year has 365 days and that the individual is susceptible at time 0. Estimate the expected values calculated in **c)** by writing computer code that simulates the Markov chain for 20 years (18250 time steps). How similar are the results achieved by simulation to the exact answers in **c)**?

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 total population $N = 1000$ 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 $\alpha = 0.01$. We 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. The discrete-time stochastic process $\{Y_n : n = 0, 1, \ldots\}$, where $Y_n = (S_n, I_n, R_n)$, is a Markov chain. We introduce 50 infected individuals in the population at time $n = 0$ so that $Y_0 = (950, 50, 0)$.

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

f) Write code that simulates the Markov chain $\{Y_n : n = 0, 1, 2, \ldots\}$ until time step $n = 300$. 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.

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`.

g) Simulate the Markov chain $\{Y_n : n = 0, 1, 2, \ldots\}$ for 100 years (36500 time steps). In the long run, approximately what proportion of individuals will be susceptible, what proportion of individuals will be infected, and what proportion of individuals will be recovered?

h) A major interest in the modelling of infectious diseases lies in the explosive behaviour during the initial breakout of the disease. Based on 1000 simulations of the outbreak for time steps $n = 0, 1, \ldots, 300$, estimate the expected maximum number of infected individuals, $\mathrm{E}[\max\{I_0, I_1, \ldots, I_{300}\}]$, and the expected time at which the number of infected individuals first takes its highest value, $\mathrm{E}[\min\{\arg \max_{n \leq 300} \{I_n\}\}]$. How would you use these two quantities to assess the potential severity of the epidemic?

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, 0:00.00.

a) Assume the rate is given by $\lambda(t) = 1.5$, $t \geq 0$. What is the probability that there are more than 100 claims before March 1 (59 days)? Verify your calculations by simulating 1000 realizations from the Poisson process. 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 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 $C_i \sim \mathrm{Exp}(\gamma)$, $i = 1, 2, \ldots$. 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 $f(c) = \gamma \mathrm{e}^{-\gamma c}$, $c > 0$. 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) Compute (by hand) the expected total claim amount and the variance of the total claim amount at March 1 (59 days) through the law of total expectation and the law of total variance. Estimate the expected value and the variance by 1000 computer simulations and compare to the true values.