TMA4265 Stochastic Modelling – Fall 2023

Project 2

Background information

- The deadline for the project is Sunday the 10th of November at 23:59.
- This project must be passed to be admitted to the final exam.
- Problem 1 and 2 are compulsory and **a reasonable attempt must be made on each of the subproblems in order to pass**.
- In addition, you need a total of **50 points** to pass. Problem 1 is worth up to 70 points, Problem 2 is worth up to 30 points, and Problem 3 is worth up to 20 points.
- Attempting Problem 3 is optional.
- The project should be done alone or with **one** group partner. You must sign up as a group in Blackboard before submitting your report and code.
- The project should be done alone or with **one** group partner. You may use the same groups you did for Project 1.
- The project report should preferably be prepared using LaTeXor `R`-markdown 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.

  The project report should be a complete text, i.e. it should be readable to a hypothetical person someone who hasn't read the problems.
- The computer code must be written in `R` of `Python` and must be submitted as a separate file **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 **8 page limit** for the project report, **or 10 pages** if you attempt the third optional problem. 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.
- We will provide help with the project in the exercise classes in week 44 and week 45.
- If you have questions outside the aforementioned times, please post your questions on Discourse (`https://mattelab2024h.math.ntnu.no/c/tma4265/75`) if possible. If your question gives away important parts of the solution, however, please instead email your question to the teaching assistant: `simen.k.furset@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: Urgent care center

Each subproblem is worth up to 10 points.

A small urgent care center (UCC) only has capacity to treat one patient at the time. Assume that the arrival of patients to the UCC follows a Poisson process with rate $\lambda > 0$, that the treatment times of the patients follow independent exponential distributions with expected value $1/\mu > 0$, and that treatment times are independent of the arrival processes to the UCC. If no one is receiving treatment, an arriving patient will immediately start receiving treatment. If another patient is currently receiving treatment, the arriving patient will join (at the end of) the queue of patients waiting for treatment. For the purpose of this project, we will assume that there is no upper limit on the number of patients that can wait inside the UCC.

Assume that $\lambda < \mu$, and let $X(t)$ denote the number of patients in the UCC at time t. We know that $X(0) = 0$.

a) Complete the following tasks:

- State the conditions for the UCC to be an M/M/1 queue, and explain why each of the conditions are satisfied.
- Explain why the stochastic process $\{X(t) : t \geq 0\}$ can be viewed as a birth-and-death process, and determine the birth rates and the death rates.
- Determine the average time a patient will spend in the UCC as a function of λ and μ.

b) Assume for this subproblem that $\lambda = 5$ patients per hour and that $1/\mu = 10$ minutes, and complete the following tasks:

- (Computer code) Simulate $\{X(t) : t \geq 0\}$ for 50 days.
- (Computer code) Based on the simulation, estimate the expected time a patient spends in the UCC. *Hint: Estimate the long-run mean number of patients in the UCC, and use Little's law to find an estimate for the expected time.*
- (Computer code) Repeat 30 times and compute an approximate 95% confidence interval (CI) for the expected time a patient spends in the UCC.
- (Report) Plot one realization of $\{X(t) : t \geq 0\}$ for the time 0–12 hours. This plot must show a series of horizontal lines (and not one continuous graph).
- (Report) Briefly describe the steps you used to estimated the expected time, give the CI, and compare the CI to the exact value computed using the expression from 1a).

The above model is unrealistic as the UCC will operate with a triage system where patients are prioritized according to urgency upon arrival. Assume that at arrival each patient will be immediately classified as urgent or normal. The probability that the patient is urgent is $0 < p < 1$, and the probability that the patient is normal is $1 - p$. We assume that the classification is independent between patients.

The UCC operates with the following queue discipline. Normal patients will not receive treatment until there are no urgent patients in the UCC. If a normal patient is receiving treatment when an urgent patient arrives, the urgent patient will immediately start receiving treatment and the normal patient will be moved back into the front of the queue.

Let $U(t)$ and $N(t)$ denote the number of urgent and normal patients, respectively, at time t. Since $X(0) = 0$, we also have that $U(0) = N(0) = 0$.

c) Complete the following tasks:

- Show that $\{U(t) : t \geq 0\}$ satifies the conditions of an M/M/1 queue.
- Determine the arrival rate.
- Determine the long-run mean number of urgent patients in the UCC as a function of λ, μ and p.

d) Complete the following tasks:

- Explain in words why $\{N(t) : t \geq 0\}$ does not behave as an M/M/1 queue.
  Hint: In one case the number of normal patients transitions from 2 to 1, and in another, it transitions from 0 to 1. How might these two cases be different?
- Determine the long-run mean number of normal patients in the UCC as a function of λ, μ and p.

e) Use Little's law to show that:

- The expected time in the UCC for an urgent patient as a function of λ, μ and p is given by
$$W_{\mathrm{U}} = \frac{1}{\mu - p\lambda}.$$
- The expected time in the UCC for a normal patient as a function of λ, μ and p is given by
$$W_{\mathrm{N}} = \frac{\mu}{(\mu - \lambda)(\mu - p\lambda)}.$$

f) Assume that $\lambda = 5$ patients per hour and that $1/\mu = 10$ minutes. We want to study how the expected time in the UCC varies as a function of the proportion of urgent patients that arrive, p. Complete the following tasks:

- Plot W_{U} and W_{N} as functions of p in the same figure.
- Give interpretations of what the situations $p \approx 0$ and $p \approx 1$ describe.
- Calculate the expected time spent at the UCC for a normal patient, W_{N}, in the extreme cases $p \approx 0$ and $p \approx 1$.
- Calculate (by hand) the p for which the expected time spent at the UCC for a normal patient is 2 hours.

g) Assume that $\lambda = 5$ patients per hour, that $1/\mu = 10$ minutes, and that $p = 0.80$. Complete the following tasks:

- (Computer code) Simulate $\{U(t) : t \geq 0\}$ and $\{N(t) : t \geq 0\}$ jointly for 50 days.
- (Computer code) Estimate the expected time in the UCC for an urgent patient, and estimate the expected time in the UCC for a normal patient. Repeat 30 times to compute an approximate 95% CI.
- (Report) Plot one joint realization of the total number of (normal and urgent) patients in the UCC and the number of urgent patients in the UCC for time 0–12 hours in the same figure. Colour them in different colours. The realizations must be displayed as series of horizontal lines (and not continuous graphs).
- (Report) Briefly describe how you estimated the two quantities, give the two CIs, and compare them to the exact numbers calculated using the expressions in 1e).

Problem 2: Calibrating climate models

Each subproblem is worth up to 10 points.

A group of climate scientists are running a climate model that outputs the temperature at every location on earth for every 6-hour period in the years 2006 and 2100[1]. The climate model is deterministic, and given the atmospheric starting conditions, external forcing, and model parameters, you will always get the same result. The challenge is that the parameters of the climate model must be selected so that the output provides as realistic evolution in time as possible. This is immensely difficult because running the model only once may require one month of computation time. For the sake of this project, assume that the only way to choose these parameters is to run the climate model for different parameter values and compare to observed temperatures.

We limit the focus to one parameter, "the albedo of sea ice", which is a measure how much sun light is reflected by sea ice. We call this parameter θ, and we decide to choose this parameter so that the temperatures observed from January 1, 2006, to October 24, 2023, matches the output of the climate model as well as possible. The fit is measured through a score $y(\theta)$ calculated based on the model output generated with parameter value θ.

The group of climate scientists have spent the last month running the model in five computing centres and provides you with five evaluation points of $(\theta, y(\theta))$: $(0.30, 0.5)$, $(0.35, 0.32)$, $(0.39, 0.40)$, $(0.41, 0.35)$, and $(0.45, 0.60)$. The observations are shown in Figure 1.

You will use a Gaussian process model $\{Y(\theta) : \theta \in [0, 1]\}$ to model the unknown relationship between the parameter value and the score. Use $\mathrm{E}[Y(\theta)] \equiv 0.5$, $\mathrm{Var}[Y(\theta)] \equiv 0.5^2$, and $\mathrm{Corr}[Y(\theta_1), Y(\theta_2)] = (1 + 15|\theta_1 - \theta_2|)\exp(-15|\theta_1 - \theta_2|)$ for $\theta_1, \theta_2 \in [0, 1]$.

a) Define a regular grid of parameter values from $\theta = 0.25$ to $\theta = 0.50$ with spacing 0.005 ($n = 51$ grid points). Construct the mean vector and the covariance matrices required to compute the conditional mean and covariance matrix of the process at the 51 grid points conditional on the five evaluation points. Display the prediction as a function of θ, along with 90% prediction intervals.

Hint: The quantiles of a Gaussian distribution can be calculated in R (`qnorm`) or Python (`scipy.stats.norm.ppf`)

b) The scientists' goal is to achieve $y(\theta) < 0.30$. Use the predictions from **a)** to compute the conditional probability that $Y(\theta) < 0.30$ given the 5 evaluation points. Plot the probability as a function of θ.

c) The scientists decide to run the climate model again with $\theta = 0.33$ and the result is $y(\theta) = 0.40$. Add this to the set of observed values, and given the six evaluation points, compute and visualize the predictions, 90% prediction intervals, and the probabilities that $Y(\theta) < 0.30$. The scientists' budget allow for one more run of the climate model, which value of θ would you suggest them to use to have the best chance to achieve $y(\theta) < 0.30$ and why?

[1] See, for example, `http://www.cesm.ucar.edu/projects/community-projects/LENS/`

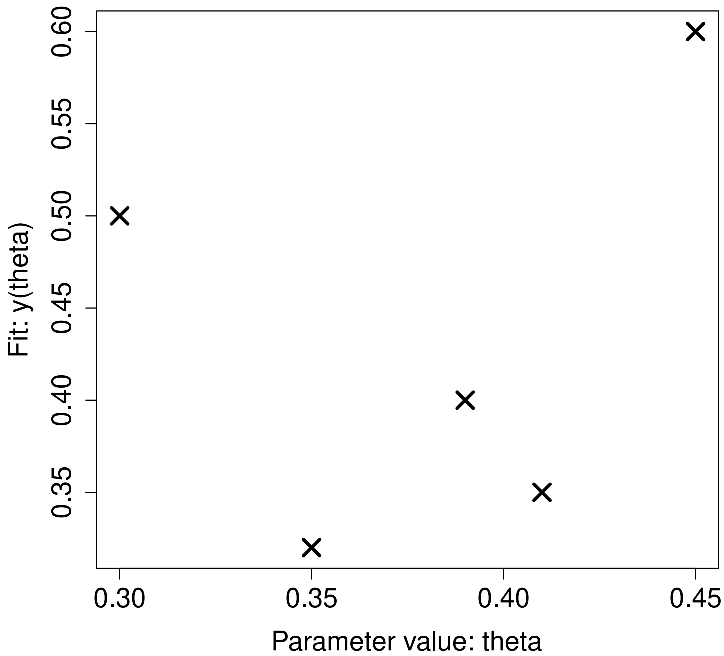

Figure 1: Observed relationship between fit and model parameter.

Problem 3: Geometric Brownian motion (OPTIONAL)

It is optional to attempt this problem, but **each subproblem is worth up to 5 points**. You may find Section 3.2 from the Gaussian process notes helpful when solving this problem, particularly in part of part c).

Let

$$X(t) = \sigma B(t), \quad t \geq 0, \tag{1}$$

where B(t) is standard Brownian motion with unit variance, $\sigma > 0$ is the scale parameter, and with $X(0) = B(0) = 0$. Let $\tau_{a,b} = \min\{t \geq 0 : X(t) \in \{-b, a\}\}$ for $a, b, > 0$ be the first time the stock reaches threshold a or $-b$. We are interested in the probability of the process reaching threshold $-b$ given it starts at 0. The probability of reaching threshold a before b is given as:

$$P(X(\tau_{a,b}) = a \mid X(0) = 0) = \frac{b}{a+b}. \tag{2}$$

The expected amount of time to first reach a or b is:

$$E[\tau_{a,b}] = \frac{ab}{\sigma^2}. \tag{3}$$

Per Ivar decided to buy some cryptocurrency for a price of 50kr and decides to sell them when they reach 75kr or drop down to 25kr. Assume the currency price, $Y(t)$, follows geometric Brownian motion, with

$$Y(t) = \exp\{\mu + \sigma B(t)\} \quad t \geq 0,$$

for $B(0) = 0$, so that $Y(0) = e^{\mu}$ and $\mu = \log(50)$, and where t is in days traded after Per Ivar bought the currency. Assume the scale parameter is given by $\sigma^2 = 4$.

a) What is the probability that Per Ivar profits from selling the currency?

b) Calculate by hand the expected time in days until Per Ivar will have to sell his cryptocurrency regardless of whether he will sell it at loss or profit from it. Write code in R to confirm the calculated values through simulation. Plot 10 realizations (simultaneously) of the process from time $t = 0$ until $Y(t)$ is either 25 or 75 kroner.

c) Numerically compute the expected time in days that Per Ivar owned the cryptocurrency before selling, conditional on the price reaching 75 kroner before 25 kroner, in the following ways:

- Using simulation in R or Python, by simulating at least 1,000 realizations of the process.
- Using numerical integration in R or Python.
  Hint: if $T > 0$ is a positive random variable, and $S(t) = P(T \geq t)$, then $E[T] = \int_0^\infty S(t)\,\mathrm{d}t$. In `R`, you may use the `integrate` function. In `Python`, you can use `scipy.integrate`. Section 3.2 from the Gaussian process notes on Blackboard is relevant here.