Stochastic Processes TMA4265

Semester project - Fall 2016

General comments (Please read them!)

- **Hand-in date: No later than 31 October, 12:00**

  Please submit **ONE document (preferably a pdf file)** which summarizes the answers to ALL questions. The document should **include all derivations, graphics and computer code**. Please be careful that the file size does not get too big.

- The project will count **20% of the final mark**.

- In order to pass this project a **reasonable attempt must be made to solve ALL problems**.

- The project should be done **in groups of two persons.**

- Computer-code should be written in `Matlab`, `R` or `Python`. Please try to make your code readable and add comments to describe what you do.

- The lecture session on 27 and 28 October will be used to answer (programming) questions.

- Questions to the project will also be answered in the exercise class at the 19 October and the 26 October. For further questions, please contact Ioannis Vardaxis (ioannis.vardaxis@math.ntnu.no).

- Deliver by email to Ioannis Vardaxis (ioannis.vardaxis@math.ntnu.no).

Useful function: There exist functions `sample, runif, rpois` in `R`, and `rand, poissrnd` in `Matlab`. Please look at the corresponding help pages to learn about their usage. For simulation of the non-homogeneous Poisson process in Exercise 3 you can use thinning. This entails drawing a homogeneous Poisson process with rate $\lambda_m = \max[\lambda(t)]$, and then independently keeping an event at time t with probability $\lambda(t)/\lambda_m$.

Exercise 1 (Discrete-time Markov chains)

The following Markov transition matrix has been used to model the air quality (0- very poor, 1 - poor, 2 - good) in a city over consecutive days:

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} 0.5 & 0.35 & 0.15 \\ 0.1 & 0.75 & 0.15 \\ 0.05 & 0.6 & 0.35 \end{pmatrix} \end{array}$$

a) Assume today's air quality is poor (state 1).
What is the probability that the air-quality will remain in the poor state for the next two days?
What is the probability that the air quality is poor the day after tomorrow?
Compute the long-run or stationary distribution of this Markov chain.

b) Use simulations to verify your answers in a):
What is the proportion of time the simulated Markov chain has air quality that is very poor? Does it agree with the theoretical answer?
Does the result depend on the initial state? And if not, show plots indicating how the initial state loses its influence?

The following Markov transition matrix has been used to model the monthly state of AIDS patients (0- dead, 1 - very sick, 2 - sick, 3 - functional):

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.05 & 0.85 & 0.1 & 0 \\ 0 & 0.05 & 0.65 & 0.3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

c) Assume a patient is very sick (state 1).
Use a first step analysis to find the probability that this patient will become functional?
Use a first step analysis to compute the expected time before this patient enters one of the absorbing states (0 or 3)?

d) Use simulations to verify your answers in c):
What is the proportion of time the simulated Markov chain ends up in the functional state? Does it agree with the theoretical answer?
Show plots indicating the variability in the stochastic process.

Exercise 2 (Markov chain Monte Carlo)

We would like to generate random realizations, i.e. draw samples, from a binomial distribution Binomial(n, p) with probability distribution

$$\pi(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x},$$

for $x = 0, 1, \ldots, n$, where n is a positive integer and $p \in [0, 1]$.

Here, we use the Metropolis-Hastings algorithm to generate a Markov chain with limiting distribution given by Binomial(n, p). Assume the present state of the Markov chain is X_{n-1} and we propose a new state $X^\star$ based on the following proposal distribution:

- If $X_{n-1} = 0$

$$Q(X^\star \mid X_{n-1} = 0) = \begin{cases} \frac{1}{2} & \text{for} \quad X^\star \in \{0, 1\} \\ 0 & \text{otherwise} \end{cases}$$

- For $X_{n-1} = k$ with $0 < k < n$

$$Q(X^\star \mid X_{n-1} = k) = \begin{cases} \frac{1}{2} & \text{for} \quad X^\star \in \{k-1, k+1\} \\ 0 & \text{otherwise} \end{cases}$$

- If $X_{n-1} = n$

$$Q(X^\star \mid X_{n-1} = n) = \begin{cases} \frac{1}{2} & \text{for} \quad X^\star \in \{n-1, n\} \\ 0 & \text{otherwise} \end{cases}$$

a) Given that the Markov chain is currently at $X_{n-1} = i$, the proposed state $X^\star = j$ will be accepted as a new state X_n of the Markov chain with probability α_{ij}. Derive the acceptance probabilities α_{ij}, for all $i \in \{0, 1, \ldots, n\}$ and $j \in \{0, 1, \ldots, n\}$.

b) Derive the transition probabilities $P_{ij} = P(X_n = j \mid X_{n-1} = i)$ of the generated Markov chain for all $i \in \{0, 1, \ldots, n\}$ and $j \in \{0, 1, \ldots, n\}$. Justify whether all assumptions are fulfilled in order that the generated Markov chain will converge to the desired distribution.

c) Write a function to implement the Metropolis-Hastings algorithm. The function should take three arguments: The parameters of the desired binomial distribution n, p and the desired length of the generated Markov chain. It should return the generated Markov chain and print the proportion of accepted proposed states.

d) Use the function implemented in part c) to simulate according to a binomial distribution with $n = 20$ and $p = 0.3$ by generating a Markov chain of length 5000. The beginning of the Markov chain is likely to not yet have converged, i.e. does not represent samples from the binomial distribution. To inspect whether beginning states have to be removed, plot X_n. If necessary remove states at the beginning (burn-in period). From the possibly reduced chain compute mean and variance and compare it to the corresponding theoretical values of the binomial distribution. Further, plot a histogram of your samples and overlay the probability distribution of the binomial distribution. What is your conclusion?

Exercise 3 (Poisson process)

Let $N(t)$ denote the number of claims received by an insurance company from time 0 to time t. Assume that $N(t)$ is a Poisson process. Here, the continuous time index $t \geq 0$ denotes days from January 1st, 0:00.00.

a) Set the intensity to a constant $\lambda(t) = 3$.
What is the probability that there are more than 200 claims before March 1st?
What is the expected waiting time of the 10th claim?
Simulate multiple Poisson processes with intensity $\lambda(t) = 3$, and use the simulations to verify your answers.
Plot the processes $N^b(t)$, for different realizations $b = 1, \ldots, 100$, as a function of time.

b) Set the intensity to $\lambda(t) = 2 + \cos(t\pi/182.5)$.
What is the probability that there are more than 200 claims before March 1st?
Simulate multiple Poisson processes with this intensity, and use the simulations to verify your answer.
Plot the processes $N^b(t)$, for different realizations $b = 1, \ldots, 100$ as a function of time.
Compare with a).

Assume that the monetary claims are independent. These claim amounts are also independent of the claim arrival times. Every claim amount (in mill. kr.) has a log-Gaussian distribution with parameters $\mu = -2$ and $\sigma^2 = 1^2$. I.e. claim amount $C_i = \exp(Y_i)$, where $Y_i \sim N(\mu, \sigma^2)$, $i = 1, 2, \ldots$.

c) Use double expectation and double variance to compute the expected total (sum of) claim amount(s) and the variance in this total claim amount when $\lambda(t) = 3$.
Simulate claim arrival times according to a Poisson distribution with constant intensity $\lambda(t) = 3$, and with $\lambda(t) = 2 + \cos(t\pi/182.5)$. Simulate claim amounts using the log-Gaussian distribution specified above.
Compare and discuss results for the two different intensities.

d) The insurance company must ensure that they can cover the total claim amount for the entire year. This must be done at the beginning of the

year, so each claim is discounted to January 1st using a discount rate $\alpha = 0.001$ (see example 5.21, page 317-19 in the book by Ross).
Compute the expected total discounted amount of claims.
How much money should the insurance company hold to be 95 percent sure they can cover the total (sum of) claims during the next year?
Do this for $\lambda(t) = 3$, and for $\lambda(t) = 2 + \cos(t\pi/182.5)$.
Use simulations to answer. (And compare with example 5.21, page 317-19 in the book by Ross.)