Stochastic Processes TMA4265

Semester project in fall 2015

General comments (Please read them!!!)

- **Hand-in date: No later than 16th of November, 12:00**

  Please submit **ONE document (preferably a pdf)** which summarises 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 has to be done **in groups of two persons.**
- Computer-code should be written in `Matlab` or `R`. Please try to make your code readable and add comments to describe what you do.
- The lecture session on the 3rd and 6th of November will be used to answer (programming) questions, and will be moved to the computer room "nullrommet 380A".
- Questions to the semester project will also be answered in the exercise class at the 28th of October and the 11th of November. (The exercise class on the 3rd of November is canceled). For further questions, please contact Ioannis Vardaxis (ioannis.vardaxis@math.ntnu.no).

All answers including derivations, computer code and graphics should be submitted via email to

Ioannis Vardaxis (ioannis.vardaxis@math.ntnu.no).

Useful function: To sample from the elements of a vector using a given probability vector, there exist the functions `sample(.)` in `R`, and `randsample` in `Matlab`. Please look at the corresponding help pages to learn about their usage.

Exercise 1 (Discrete-time Markov chains)

Assume we are interested in modelling the status of one airplane. Each time the airplane lands we classify its status as: 0 - early, 1 - on-time, 2 - delayed. We use a Markov chain to model the status and assume the following transition probability matrix

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} 0.1 & 0.85 & 0.05 \\ 0.1 & 0.75 & 0.15 \\ 0.05 & 0.6 & 0.35 \end{pmatrix} \end{array}$$

Assume the airplane is currently in state 1.

a) Find the expected number of flights until the plane is no longer on-time. Compare your theoretical result to the one obtained when simulating the process.

b) What is the expected number of flights until the plane is the first time delayed. Compare your theoretical result to the one obtained by simulations

c) Simulate a sequence of 1000 flights. What is the proportion of time the plane is delayed? Does it agree with the theoretical answer? Does the result depend on the initial state? Justify your answer.

Exercise 2 (Markov chain Monte Carlo)

We would like to generate random realisations, 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 $n \in \mathbb{N}_0$ 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 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 our desired distribution.

c) Write a function to implement the Metropolis-Hasting 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 (Continuous-time Markov chains)

Let $X(t)$ denote the number of individuals at time t in a linear growth model with immigration, in which

$$\mu_n = n\mu \quad , n \geq 1$$
$$\lambda_n = n\lambda + \theta \quad , n \geq 0.$$

Each individual in the population is assumed to give birth at an exponential rate λ. Further, there is an exponential rate of increase θ of the population due to an external source such as immigration. Hence, the total birth rate when there are n persons in the system is $n\lambda + \theta$. Deaths are assumed to occur at an exponential rate μ for each member of the population. Assume there are i individuals in the system at $t = 0$, i.e $X(0) = i$.

From the lecture we know:

$$E[X(t)] = \frac{\theta}{\lambda - \mu}[e^{(\lambda-\mu)t} - 1] + ie^{(\lambda-\mu)t}$$

if $\lambda \neq \mu$. If $\lambda = \mu$ we obtained

$$E[X(t)] = \theta t + i.$$

Now we would like to simulate trajectories, i.e. sample paths, of $X(t)$.

a) Describe an algorithm that will generate sample paths of $X(t)$ over time $[0, T]$ and implement the algorithm as a function. The function should take five arguments: The initial population size i, μ, λ, θ and T. The algorithm should return a matrix with two columns: the time points t at which events are observed and the corresponding state $X(t)$.

b) Plot 50 simulated trajectories in one figure for

   i) $X(0) = 50$, $\lambda = 0.02$, $\mu = 0.02$, $\theta = 0.02$, $T = 50$. What happens if you increase θ?

   ii) $X(0) = 2$, $\lambda = 0.08$, $\mu = 0.04$, $\theta = 1$, $T = 50$. How does the process change when increasing the initial value i? Consider also the variability of the process in your description.

   iii) What happens if you switch the values of μ and λ in part ii)? What if $\theta = 0$?

   Compare also in each case $E[X(t)]$ with the mean estimated from the 50 sample paths by including both the theoretical and estimated expected value as lines in your plot.

c) Let $X(0) = 50$, $\lambda = 0.06$, $\mu = 0.02$, $\theta = 0.02$. When do you expect to have the population doubled for the first time? Compare the theoretical result with the estimate obtained based on 100 simulations. Based on the simulations determine how certain you are about this prediction in terms of the standard deviation. (Hint: In `R` the function `match` might be useful)