MVE550 2020 Lecture 10.1
Dobrow Sections 7.1, 7.2, 7.3
Continuous-time Markov chains

Petter Mostad

Chalmers University

November 27, 2020

Introduction

- We now consider general continuous-time discrete state space Markov chains.
- Comparing to the discrete-time Markov chains we studied before: We now model that we stay in each state for some real-valued amount of time.
- The Markov property is a type of "memoryless-ness": The property will imply that the amount of time in each state is Exponentially distributed.
- Very useful tool, can be used to model for example queues.

Example

- We have previously discussed modelling the weather as a discrete time Markov chain where the weather *each day* is "rain", "snow", or "clear", with transition matrix for example

$$P = \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.6 & 0.3 \end{bmatrix}.$$

- A more realistic model is that each weather type lasts some length of time, before changing to a *different* weather type:
  - Let's say the time each weather type lasts is Exponentially distributed with parameters q_r, q_s and q_c (so that expected durations of weather types are $1/q_r$, $1/q_s$, $1/q_c$, respectively).
  - Transitions after this time would happen according to a transition matrix

$$\tilde{P} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 3/4 & 0 & 1/4 \\ 1/4 & 3/4 & 0 \end{bmatrix}.$$

- Note that the process is completely described by parameters q_r, q_s, q_c and p_{ij}, where $\tilde{P}_{ij} = p_{ij}$. Note that $p_{ii} = 0$ for all i.

Continuous time Markov chains

- A continuous time stochastic process $\{X_t\}_{t\geq 0}$ with discrete state space S is a *continuous time Markov chain* if

$$P(X_{t+s} = j \mid X_s = i, X_u, 0 \leq u < s) = P(X_{t+s} = j \mid X_s = i)$$

where $s, t \geq 0$ and $i, j, x_u \in S$.

- The process is *time-homogeneous* if for $s, t \geq 0$ and all $i, j \in S$

$$P(X_{t+s} = j \mid X_s = i) = P(X_t = j \mid X_0 = i)$$

.

- We then define the *transition function* as the matrix function $P(t)$ with the entries of the matrix given by

$$P(t)_{ij} = P(X_t = j \mid X_0 = i)$$

The Chapman-Kolmogorov Equations

For the transition function $P(t)$ we have

- $P(s+t) = P(s)P(t)$
- $P(0) = I$
- Note similarity to the properties of the exponential function! However, $P(t)$ is a matrix, not a number.
- Example:
  - A Poisson process with parameter λ is a continuous time time-homogeneous Markov chain.
  - We get

$$P(t) = \begin{bmatrix} e^{-\lambda t} & (\lambda t)e^{-\lambda t} & (\lambda t)^2 e^{-\lambda t}/2! & (\lambda t)^3 e^{-\lambda t}/3! & \dots \\ 0 & e^{-\lambda t} & (\lambda t)e^{-\lambda t} & (\lambda t)^2 e^{-\lambda t}/2! & \dots \\ 0 & 0 & e^{-\lambda t} & (\lambda t)e^{-\lambda t} & \dots \\ 0 & 0 & 0 & e^{-\lambda t} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Holding times are exponentially distributed

- Define T_i as the time the continuous-time Markov chain started in i stays in i before moving to a different state, so that for any $s > 0$

$$P(T_i > s) = P(X_u = i, 0 \leq u \leq s)$$

- The distribution of T_i is *memoryless* and thus exponential.
- We define q_i so that

$$T_i \sim \text{Exponential}(q_i)$$

- Remember that this means that the average time the process stays in i is $1/q_i$. The *rate* of transition out of the state is q_i.
- Note that we can have $q_i = 0$ meaning that the state i is *absorbing*: $P(T_i > s) = 1$.

The embedded chain

- Define a new stochastic process by listing the states the chain visits. This will be a discrete time Markov chain.
- It is called the *embedded chain*; transition matrix is denoted $\tilde{P}$.
- Note that $\tilde{P}$ has zeros along its diagonal!
- Note that the continuous time Markov chain is completely determined by the expected holding times $(1/q_1, \ldots, 1/q_k)$ and the transition matrix $\tilde{P}$ of the embedded chain. We write p_{ij} for its entries.

Describing the chain using transition rates

A way to describe a continuous-time Markov chain is to describe $k \times (k-1)$ independent "alarm clocks":

- For states i and j so that $i \neq j$, let q_{ij} be the parameter of an Exponentially distributed random variable representing the time until an "alarm clock" rings.
- When in state i, wait until the first alarm clock rings, then move to the state given by the index j of that alarm clock. This defines a continuous-time Markov chain.
- The time until the first alarmclock rings is Exponentially distributed with parameter given by

$$q_i = q_{i1} + q_{i2} + \cdots + q_{i,i-1} + q_{i,i+1} + \cdots + q_{ik} \tag{1}$$

  i.e., the parameter of the holding time distribution at i.
- We will se below: The chain is completely described by the rates q_{ij}, $i \neq j$.
- We saw above: The chain is also completely determined by the p_{ij} and the q_i. The relationship is Equation 1 and, for $i \neq j$,

$$p_{ij} \cdot q_i = q_{ij}.$$

The derivative of $P(t)$ at zero

- To relate $P(t)$ to the q_{ij}'s, we first relate them to $P'(0)$.
- Assuming $P(t)$ is differentiable we get

$$P'(0) = \begin{bmatrix} -q_1 & q_{12} & q_{13} & \cdots & q_{1k} \\ q_{21} & -q_2 & q_{23} & \cdots & q_{2k} \\ q_{31} & q_{31} & -q_3 & \cdots & q_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{k1} & q_{k2} & q_{k3} & \cdots & -q_k \end{bmatrix} = Q$$

  where the q_i and the q_{ij} are those defined earlier.
- In fact we don't need to require a finite state space; discrete is enough.
- Q is called the *(infinitesimal) generator* of the chain.

Kolmogorov Forward Backward

- Prove: We get that for all $t \geq 0$,

$$P'(t) = P(t)Q = QP(t)$$

- Note what this means in terms of the components of $P(t)$:

$$\begin{aligned} P'(t)_{ij} &= -P_{ij}(t)q_j + \sum_{k \neq j} P_{ik}(t)q_{kj} \\ P'(t)_{ij} &= -q_i P_{ij}(t) + \sum_{k \neq i} q_{ik} P_{kj}(t) \end{aligned}$$

- The equations above define a set of differential equations which the components of the matrix function $P(t)$ needs to fulfill.

MVE550 2020 Lecture 10.2
Dobrow Sections 7.3, 7.4
Matrix exponential. Limiting behaviour.

Petter Mostad

Chalmers University

November 28, 2020

Review / motivation

For continuous time time-homogeneous Markov chains with discrete state space:

- Let $P(t)$ be the matrix of probabilities for changing from state i to state j after time t.
- We found that

$$P'(0) = \begin{bmatrix} -q_1 & q_{12} & q_{13} & \cdots & q_{1k} \\ q_{21} & -q_2 & q_{23} & \cdots & q_{2k} \\ q_{31} & q_{31} & -q_3 & \cdots & q_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q_{k1} & q_{k2} & q_{k3} & \cdots & -q_k \end{bmatrix} = Q$$

  where the q_{ij} are ("alarm clock") rates and the rows of Q sum to 0. We know these rates, i.e., Q, determine the whole process!
- We found that $P'(t) = QP(t) = P(t)Q$.
- It seems tempting to define $P(t) = e^{tQ}$. But can we do that when Q is a matrix?

The matrix exponential

- For any square matrix A define the *matrix exponential* as

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n = I + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \frac{1}{24}A^4 + \dots$$

- The series converges for all square matrices A (we don't show this).
- Some important properties:
  - $e^0 = I$.
  - $e^A e^{-A} = I$.
  - $e^{(s+t)A} = e^{sA} e^{tA}$.
  - If $AB = BA$ then $e^{A+B} = e^A e^B = e^B e^A$.
  - $\frac{\partial}{\partial t} e^{tA} = A e^{tA} = e^{tA} A$.
- $P(t) = e^{tQ}$ is the unique solution to the differential equations $P'(t) = QP(t)$ for all $t \geq 0$ and $P(0) = I$.
- In `R` you may use `expm` from R package `expm` to compute exponential matrices.

Computing the matrix exponential

- Assume there exists an invertible matrix S and a matrix D such that $Q = SDS^{-1}$. Then (show!)

$$e^{tQ} = Se^{tD}S^{-1}$$

- If $D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_k \end{bmatrix}$ is a diagonal matrix, then (show!)

$$e^{tD} = \begin{bmatrix} e^{t\lambda_1} & 0 & \dots & 0 \\ 0 & e^{t\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{t\lambda_k} \end{bmatrix}.$$

- Recall that if Q is *diagonalizable* it can be written as $Q = SDS^{-1}$ where D is diagonal with the eigenvalues along the diagonal, and S has the corresponding eigenvectors as columns.

Limiting and stationary distributions

- A probability vector v represents a *limiting distribution* if, for all states i and j,
$$\lim_{t \to \infty} P_{ij}(t) = v_j.$$
- A probability vector v represents a *stationary distribution*, if, for all $t \geq 0$,
$$v = vP(t)$$
- Note: This happens if and only if $0 = vQ$.
- A limiting distribution is a stationary distribution but not necessarily vice versa.
- A continuous-time Markov chain is *irreducible* if for all i and j there exists a $t > 0$ such that $P_{ij}(t) > 0$.
- However, periodic continuous-time Markov chains do not exist: If $P_{ij}(t) > 0$ for some $t > 0$ then $P_{ij}(t) > 0$ for all $t > 0$.

The fundamental limit theorem

- An absorbing communication class is one where there is zero probability (i.e., zero rate) of leaving it to other commuication classes.
- For a finite-state continuous-time Markov chain with finite holding time parameters, there are two possibilities:
  - The process is irreducible, and $P_{ij}(t) > 0$ for all $t > 0$ and all i, j.
  - The process contains one or more absorbing communication classes.
- **Fundamental Limit Theorem**: Let $\{X_t\}_{t \geq 0}$ be a finite, irreducible, continuous-time Markov chain with transition funciton $P(t)$. Then there exists a unique stationary distribution vector v which is also the limiting distribution.
- The limiting distribution of such a chain can be found as the unique v satisfying $vQ = 0$.

Stationary distribution of the embedded chain

- Recall the *embedded chain* of a continuous-time Markov chain.
- Stationary distributions for the embedded chain and for the continuous-time chain are generally not the same!
- However, there is a simple relationship: A probability vector π is a stationary distribution for a continuous-time Markov chain if and only if ψ is a stationary distribution for the embedded chain, where $\psi_j = C\pi_j q_j$ for the appropriate constant C.