Bioinformatics Research Center
Aarhus University

Statistical Methods in Bioinformatics
Asger Hobolth
November 19, 2018

Handout 12

Lectures in Week 47

EG Section 11.7: Fundamental mathematical and statistical properties for continuous time Markov chains.

Lectures in Week 48

EG Section 14.3: Applications of continuous time Markov chains in molecular evolution.

Exercises in Week 47

1. **The symmetric 2-state chain**
   EG Problem 11.10 page 407.

2. **The symmetric 2-state chain**
   Let
   $$Q = \begin{pmatrix} -\alpha & \alpha \\ \alpha & -\alpha \end{pmatrix}.$$
   Use induction (see EG B.18 page 556) to show that for $i = 1, 2, \ldots$ we have
   $$Q^i = \frac{1}{2}(-2\alpha)^i \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$
   Use the Taylor series expansion for the exponential function (see EG eqn. (B.20) page 551) to show
   $$\sum_{i=1}^{\infty} \frac{t^i}{i!} Q^i = \frac{1}{2}\big(\exp(-2\alpha t) - 1\big) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$
   Determine $e^{Qt} = \sum_{i=0}^{\infty}(t^i/i!)Q^i$, and relate your result to the previous exercise (EG Problem 11.10).

3. **The general 2-state chain**
   Consider the general 2-state continuous time Markov chain with rate matrix
   $$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix},$$
   where states are numbered 0,1.

   (i) Show that the Kolmogorov Forward Equation becomes
   $$P_{00}^{'}(t) = \mu P_{01}(t) - \lambda P_{00}(t).$$

(ii) Use $P_{01}(t) = 1 - P_{00}(t)$ and $P_{00}(0) = 1$ to show that

$$P_{00}(t) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)t}.$$

4. **EG Section 14.3.1: The Jukes-Cantor model**
   Show the transition probabilities EG equation (14.27) and (14.28) for the Jukes-Cantor model: Repeat the arguments on top of EG page 485.

5. **Simulation of a Continuous Time Markov Chain**

   (a) **Forward simulation.** Write a program in R that simulates from a CTMC. Input should be a rate matrix Q, an initial distribution ϕ and the length of the time interval T. Output should be a sample path from the stochastic process.

   (b) **Rejection sampling.** Write a program in R that simulates a sample path from a CTMC with rate matrix Q, beginning state a, ending state b and time interval T.
   Hint: One way of accomplishing the task is to simulate forward like in (a), and accept the sample path when the ending state is b, and reject otherwise.

   (c) **Conditional means.** Write a program i R that can estimate the conditional means

   $$\mathrm{E}_Q[N(i,j)|X(0) = a, X(T) = b] \text{ and } \mathrm{E}_Q[T(i)|X(0) = a, X(T) = b],$$

   where $N(i,j)$ is the number of jumps from i to j and $T(i)$ is the time spent in state i. Here $\{X(T) : t \geq 0\}$ is the CTMC with rate matrix Q.