Markovian modelling and Bayesian learning, fall 2014

Lecturer
Jukka Corander

Scope
5 sp.

Type
Advanced studies

Prerequisites
Basic calculus, linear algebra, introductory course on probability and statistical inference are absolutely necessary. First course level knowledge on algebra, probability and inference will be recommendable for many parts of the course. Short course synopsis:

Part 1. Basic properties of discrete-time Markov chains (DTMCs), irreducibility, ergodicity, stationarity, invariant distributions, higher order Markov chains

Part 2. Statistical inference for DTMCs, maximum likelihood estimation, Bayesian estimation, inference about the order of a DTMC, full likelihood, model averaging, applications to clustering of DNA sequences

Part 3. Continuous-time Markov chains (CTMCs), basic properties, waiting time distributions, matrix forward equations, generator, absolute vs relative time, maximum likelihood estimation, unidentifiability of root position, applications to DNA and amino acid sequence evolution

Part 4. Hidden Markov models (HMMs). Basic properties, inference tasks related to smoothing, filtering and prediction, Derin's algorithm, factorization and recursion for HMM calculations, applications to classification and modeling of DNA sequences

Part 5. Variable-order DTMCs (VOMs/VLMCs), sparse higher-order Markov chains (SMCs), inference for VOM/VLMC/SMC models, applications to sequence prediction and classification

Lectures
Weeks 44-49, Tuesday 12-14 and Thursday 12-14 in room B120. Course end on December 4th.

Exercises
During weeks 45-49 there will be a weekly exercise session on Thursdays 14-16. On Nov 13 it is in D122 and in C124 for last three sessions. The teachers responsible for the exercise sessions are Elina Numminen (weeks 45-46) and Jie Xiong (weeks 47-49), both have 'at' helsinki dot fi address. Solutions to each week's exercises must be presented in the session or at least 1 hour beforehand by email to the responsible teacher.

Exercises for week 45 are available here
Exercises for week 46 are available here
Exercises for week 47 are available here
Exercises for week 48 are available here
Exercises for week 49 are available here

Exams
To gain the credits from this course, it is necessary to do at least 50% of the exercises and a home exam. Additional solved exercises will yield bonus points for the grade. The home exam will consist of a number of larger assignments that must be returned by May 1st 2015 to the lecturer. Home exam assignments are available here.

Preliminary lecture diary
Three examples on why Markovian type models are useful: modeling the current Ebola outbreak, HMM and tracking of objects (examples starting at 49:30), compression of DNA sequence data (backup of the paper if don't have access rights), some slides about modeling neutral evolution, Teaser trailer, Eye-opener on conditional probabilities and Bayes' theorem, basic properties of Markov chains. This excerpt from the HMM book by T. Koski is mainly used during the lecture and also this short excerpt on periodicity from the book of Isaacson & Madsen, Markov chains. For further illustrations and mathematical details on Markov chains, see the link to Sirl and Norris in Bibliography.

Basic properties of Markov chains continued. To get going with the basics of simulating Markov chains, you might find these Matlab codes useful.

Bayesian estimation of DTMC parameters, Continuous-time Markov chains (CTMCs) continued, for an application of birth-death processes to modeling the 2014 Ebola epidemic see this paper, here is also an application of CTMCs to infectious disease epidemiology.

A primer on Occham's razor and Bayesian model comparison for Markov chains, Information-theoretic book by D MacKay where Ch 28 contains a detailed explanation of the Occam's razor principle and Bayesian model comparison, Bayesian learning of the order of a DTMC, see also the use of information-theoretic criteria for choosing the model dimension as explained in this nice review paper, use of Markov chains to clustering of DNA sequences in metagenomics applications, paper1 (published in the journal Statistical Applications in Genetics and Molecular Biology), paper2

Bibliography

Various references will be used during the course. The lecture diary will also include links to some additional materials. Parts of the following books will be considered:


In addition, we will consider a number of articles & tutorials (articles not directly linked here are generally available form JSTOR collection or are otherwise online):

Norris, J. Markov chains. CUP, see online resource.
Gu, L. Notes on Dirichlet distribution with relatives. This document provides a concise recapitulation of some of the central formulas that are needed in the exercises and assignments when doing Bayesian learning. More comprehensive derivations can be found in several books on Bayesian modeling, e.g. in Koski & Noble (2009), which is listed above.

Registration

Did you forget to register? What to do?