

Computational statistics, fall 2016

Computational statistics, fall 2016

Teacher: Christian Benner (christian.benner at helsinki.fi)

Scope: 5 cr (one period) or 10 cr (two periods)

Type: Advanced studies in statistics. Compulsory course in Statistical Machine Learning degree requirements. It is also possible to include the course as intermediate level statistics course.

Teaching: Lectures, exercises and computer class working in class C128, Kumpula campus, Exactum

Teaching schedule:

- I teaching period, 5 cr: 10-12 in Mon, Wed, Fri, 5.-16.9 and Mon, Wed, 19.9 - 12.10
- II teaching period, intensive course, 5 cr: 9-12 in Mon, Tue, Wed, Thu, Fri, 7.-11.11 and Thu 17.11, Mon 21.11 as well as Wed 23.11

Topics:

- The I-period part of the course gives an overview of computational methods which are useful especially in Bayesian statistics (but some of the methods are also used widely in frequentist inference).
Background handbook:
 - [Petri Koistinen, Computational statistics. 2013. Chapter 1-4.](#)
 - [Petri Koistinen, Computational statistics. 2013. Chapter 5-6.](#)
 - [Petri Koistinen, Computational statistics. 2013. Chapter 7-11.](#)
- The II-period intensive part of the course is about implementing a computational method for statistical variable selection. This part is very hands-on and will live from discussions and feedback during the sessions.

Prerequisites: Courses 57703 Data-analysis with R, as well as compulsory intermediate level statistics courses (57705 Probability calculus II, 57701 Statistical inference II, 57714 Linear models I) are prerequisites for this course. The compulsory (in math-stat department statistics degree requirements) intermediate level course 57753 Bayesian inference is not a prerequisite.

News

- The website will be continuously updated from now on.
- We will have to change the times 14.-17.11 for II teaching period because I'll be abroad.
- On Wednesday Sep 20 we finished conjugate analysis and looked at Gibbs sampling.

Exercises/Assignment

- [Exercise 0 \(solutions\)](#)
- [Exercise 1 \(solutions\)](#)
- [Exercise 2](#)
- [Exercise 3](#)
- [Exercise 4](#)
- [Exercise 5](#)
- [Assignment](#) (deadline is Dec 9, 2016)

Exercises are to be solved before each session. The solutions and their implementation as well as particular theory concepts will be discussed during each session. You will get additional points from solving exercises. These points will be added to your points from course exams, according to the formula $\max(0, \text{floor}((n - 2) / 5))$. There will be a list going around during each session.

R code

- Predictive density
- Inverse transform sampling
- Accept-reject sampling
- [Gibbs sampling and mean-field VB approximation](#)
- [Matched curvature candidate density in independent Metropolis-Hastings sampling](#)

Exams

The exam for I-period part will be on Oct 24, 2016 at 10:00 in D122.

Home assignment II-period part

The home assignment (2-3 pages, not including title page) documents together with your C++ code how you solved the problem. Due date will be decided on.

Course material

I-period part

There will be several examples which show how the methods can be implemented using the R system for statistical computing. R is convenient for us since it is freely available and widely used and it enables easy visualization of results and contains simulation functions for lots of distributions. However, the methods are in no way tied to the R environment, and the methods can as easily be used in many other environments (such as Matlab together with its statistics toolbox).

Background handbook:

- Petri Koistinen, Computational statistics. 2013. Chapter 1-4.
- Petri Koistinen, Computational statistics. 2013. Chapter 5-6.
- Petri Koistinen, Computational statistics. 2013. Chapter 7-11.

Topics:

- Review of probability and Bayesian inference
- Methods for generating independent samples from distributions
- Classical Monte Carlo integration and importance sampling
- Approximating the posterior distribution using numerical quadrature or Laplace expansion
- MCMC methods: Gibbs and Metropolis-Hastings sampling
- Auxiliary variable methods in MCMC
- EM algorithm
- Multi-model inference
- MCMC theory

II-period part

This part of the course is about implementing a computational method for statistical variable selection. The implementation will be carried out within the C++ programming language but **you do not need to know C++ as a prerequisites for participating in the course**. Several examples and data-sets will be used to illustrate the underlying methodology and get you started in C++.

When you register to the course, you don't have to decide whether you take the 5 cr option or 5+5 cr option.

Registration

Did you forget to register? [What to do?](#)

Course feedback

Course feedback can be given at any point during the course. Click [here](#).