

Computational statistics, fall 2014

Computational statistics, fall 2014

Lecturer

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

Type

Advanced studies in statistics

Scope

5-10 cr. It is possible to take only the I-period part (5 cr) or I+II periods (10 cr)

Prerequisites

Courses 57703 Data-analysis with R and 57753 Bayesian inference, as well as all compulsory intermediate level statistics courses (57705, 57701, 57714) are prerequisites for this course

Lectures / Computer class sessions

Weeks 36-42 (I-period part) and 44-50 (II-period part), Wednesday 12-16 in computer class C128

Content & Exercises / I-period part

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)

- 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.

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).

References to chapters apply to the text by Petri Koistinen.

- [Exercise set 0](#) (03.09.2014)
- [Exercise set 1](#) (10.09.2014)
- [Exercise set 2](#) (17.09.2014)
- [Exercise set 3](#) (24.09.2014)
- [Exercise set 4](#) (01.10.2014)
- [Exercise set 5](#) (08.10.2014)

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 and being able to present them. 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.

Content / II-period part

The II-period part of the course is about implementing some computational intensive statistical method. The implementation will be carried out within the C++ programming language using high-performance linear algebra libraries and other parallel and GPU computing approaches. Several examples and data-sets will be used to illustrate the methods.

Readings:

- [Zhaojun Bai and Gene Golub. Bounds for the Trace of the Inverse and the Determinant of Symmetric Positive Definite Matrices](#)
- [Gérard Meurant. Estimates of the trace of the inverse of a symmetric matrix using the modified Chebyshev algorithm](#)
- [Zhaojun Bai, Mark Fahey and Gene Golub. Some large-scale matrix computation problems](#)

Assignment:

- [Until 28.02.2015](#)

Exams / Home assignment

I-period part of the course ends with a course exam and [home assignment](#) (18.01.2015). The course exam will be in week 42, Wednesday 15.10.2014 in C128 at 12:00.

II-period part of the course ends with a home programming project and short report that describes the statistical model and method.

Bibliography

Petri Koistinen, Computational statistics. 2013. Chapter 1-4.

Petri Koistinen, Computational statistics. 2013. Chapter 5-6.

Petri Koistinen, Computational statistics. 2013. Chapter 7-11.

Dimitri Bertsekas and John Tsitsiklis. Introduction to Probability. Nashua (NH): Athena Scientific; 2002.

William Bolstad. Understanding computational Bayesian statistics. Hoboken (NJ): John Wiley & Sons; 2010.

Geof Givens and Jennifer Hoeting. Computational statistics. Hoboken (NJ): John Wiley & Sons; 2012.

Registration

Registration is to 5-10 cr, i.e. when you register, you don't have know, whether you continue to 5 cr -> 10 cr. There will not be a separate registration to -> 10 cr part.

Did you forget to register? What to do?