

# Computational statistics, fall 2015

## Computational statistics, fall 2015

**Teacher:** Christian Benner (email: christian dot benner at helsinki dot fi)

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

**Type:** Advanced studies in statistics. Compulsory course in Statistical Machine Learning degree requirements.

**Teaching:** Lectures, exercises and computer class work.

**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). See below for more details.  
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 part of the course is about implementing a computational method for fine-mapping results from Genome-Wide Association Studies (GWAS). See below for more details.

**Prerequisites:** Courses 57703 Data-analysis with R, as well as compulsory intermediate level statistics courses (57705, 57701, 57714) 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

- We start Monday 14.09.2015 (week 38) with a review of probability theory and Bayesian inference. The session from week 36 will take place during week 38-42.
- There is no session in week 41 because I will be outside of Finland.
- **The canceled sessions will be replaced with sessions on Friday 30.10.2015 (week 44) and 06.11.2015 (week 45). Both sessions take place from 12-16 in room B120.**

## Teaching schedule

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

## Exercises/Assignment

- [Exercise 0 \(solutions\)](#)
- [Exercise 1 \(solutions\)](#)
- [Exercise 2 \(solutions\)](#)
- [Exercise 3 \(solutions\)](#)
- [Exercise 4 \(solutions\)](#)
- [Exercise 5 \(solutions\)](#)
- Assignment (due Sunday, January 31, 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
- Importance sampling

## Exams

The exam will be on December 15 from 10-14 at FIMM. The address is Tukholmankatu 8 (Biomedicum Helsinki 2U). We meet at the entrance (<http://www.helsinki.fi/teknos/opetustilat/meilahti/t8u/default.htm>).

# 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 is February 29, 2016.

## 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 fine-mapping results from Genome-Wide Association Studies (GWAS). For an excellent review on fine-mapping, see Spain et al. (2015). The implementation will be carried out within the C++ programming language. Several examples and data-sets will be used to illustrate the underlying methodology and get you started in C++.

Complementary course in period II:

- 57746 Genome-wide association studies by Matti Pirinen (FIMM)

## References

- Spain, S. and Barrett, J. (2015) Strategies for fine mapping complex traits, Hum. Mol. Genet., 42, 1001-1006.
- Benner, C. et al. (2015) FINEMAP: Efficient variable selection using summary data from genome-wide association studies, bioRxiv doi: <http://dx.doi.org/10.1101/027342>

## Registration

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

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

## Course feedback

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