

# M5MS04 – Computational Statistics

MSc Statistics – Core Course  
Autumn 2013

## Instructors:

|                          |                |                                  |
|--------------------------|----------------|----------------------------------|
| Professor David Stephens | 544 Huxley     | <i>d.stephens@imperial.ac.uk</i> |
| Dr Daniel Mortlock       | 1017A Blackett | <i>d.mortlock@imperial.ac.uk</i> |

## Class Meetings:

Mondays 14.00–16.00 and Wednesdays 13.00–14.00 in Huxley 139

|                      |          |           |             |
|----------------------|----------|-----------|-------------|
| <b>Office Hours:</b> | Stephens | Wednesday | 14.00-15.30 |
|                      | Mortlock | Wednesday | 14.00-15.00 |

Objectives: Modern applied statistics is heavily dependent on sophisticated computational methods. Indeed, fitting the multi-level and highly hierarchical models that are becoming ever more prevalent in the biological, engineering, physical, and social sciences would not be feasible without the recent rapid advances in computing. This course will develop a basic tool kit of computational methods useful in applied statistics. Specific topics include optimization, quadrature, Monte Carlo methods, and numerical linear algebra.

Prerequisites: Students are assumed to have a solid background in probability and statistical inference equivalent to an undergraduate degree in statistics.

Lectures: Class will meet for three hours each week. Often, one hour each week will be devoted to computer demonstrations and hands-on computer experiments, while more formal lectures will occupy the remainder of the class time. Students should bring a laptop computer to lectures in order to participate in the demonstrations.

Textbooks: Course material will be drawn from:

**Primary Text:** Givens, G., and Hoeting, J. (2005). *Computational Statistics*, Wiley.

Brooks, S., Gelman, A., Jones, G., and Meng, X. L. (2011). *Handbook of Markov Chain Monte Carlo*, Chapman & Hall.

Lange, K. (2010). *Numerical Analysis for Statisticians, 2nd Edition*, Springer-Verlag.

Computing: Homework and class projects will require significant statistical computing. We will be using the software package R. This package is freely available: URL: [www.r-project.org](http://www.r-project.org). R-programming will be discussed throughout the course as needed to facilitate the computing topics. Students are assumed to have completed the Departmental three-hour “crash course” in R, but no other experience with R is assumed.

### Outline of Topics:

1. Univariate Optimization Methods: optimization methods including bisection, functional iteration, and Newton's method; topics in optimization theory including linear and quadratic rates of convergence and local and global convergence.
2. Multivariate Optimization Methods: optimization methods including Newton-type methods, steepest ascent, Fisher scoring, iteratively reweighted least squares, Gauss-Seidel algorithm, EM algorithms, and EM-type algorithms; topics in optimization theory including assessing convergence, stability and speed, improving stability, and accelerating convergence.
3. Simulation via Direct Methods: Random number generation including uniform distributions and other distributions; general simulation methods including inverse cdf generation, the grid method, and rejection sampling.
4. Numerical Integration: Laplace's method; Gaussian quadrature; Monte Carlo integration; Blackwellization; importance sampling; nested sampling; the VEGAS algorithm.
5. Simulation via Markov Chains: MCMC methods such as Metropolis-Hastings, Gibbs sampler, data augmentation, and advanced MCMC methods. MCMC topics including efficiency of Markov chain algorithms, assessing convergence, rates of convergence, improving convergence, and decorrelation methods.
6. Additional Topics (time permitting): bootstrap, Bayesian bootstrap, parametric bootstrap, posterior predictive p-values, and cross-validation; adaptive rejection sampling; reversible jump methods; auxiliary variables and the slice sampler; simulated tempering and annealing; perfect simulation; bridge sampling; computing Bayes factors; sequential Monte Carlo.

### Course requirements:

There will be several homework assignments during the term. The final project will be an application of some of the methods developed in the course. There will be no exams in this course.

### Grades:

|                      |     |
|----------------------|-----|
| Homework assignments | 50% |
| Final project        | 50% |