Course Outline

Summer Reading

Please go through the MSc R programming notes and practicals on Robin Evans' webpage before the course starts:

`http://www.stats.ox.ac.uk/~evans/Rprog/index.htm`

We assume that you are either happy and comfortable with the material in the notes or have a list of things you would like better explained when you arrive.

Abstract

Computational statistics is one of the most dynamic areas of research in modern statistics and is key to such fields as Bayesian inference, model choice, graphical models, Big Data and others. In this course, students will be exposed to main concepts of computational statistics and to milestone ideas, such as Monte Carlo, Markov chain Monte Carlo, Sequential Monte Carlo, Reversible Jump, Approximate Bayesian Computation, Intractable Likelihood, Monte Carlo Expectation Maximisation, Simulated Annealing, Belief Propagation, LASSO, LARS, false discovery rate, to name a few, and will study two of them in depth based on original research papers. The study will involve implementing one of the methodologies and may involve analysis of real data. The course is designed not only to cover computational statistics and R, but also to support acquisition of transferable skills such as collaborative work, giving and receiving constructive feedback and good practice in producing code.

Timetable

- **Mon 9th October:** Computational problems in statistics - an introduction; Review of R; including calling C code, debugging; some practicals.
- **Tue:** Presentation of $6-8$ classic papers from computational statistics and related projects, about 30 minutes per topic.
- **Tue:** Evening: breaking into pairs; assigning projects to the pairs (each pair will have one Oxford and one Warwick student)
- **rest of week 1:** work on the project: produce code and draft reports
- **Mon 16th October: peer review day:**
  9am: hand in the theory part of the report to another team (not reciprocal) and email to the module leaders (this draft version will not be assesed by module leaders, but submission is compulsory);
  2pm: review second team's work, give feedback (round 1)
  4pm: review second team's work, give feedback (round 2)
- **Tue:** project work +
  pizza-lunch + lunch time speaker (from Oxford or Warwick)
- **Wed:** finalise your project
  6pm: hand in code (in the form of an R vignette) and reports for marking

- **Thu:** work on presentation +
pizza-lunch + lunch time speaker (from Oxford or Warwick)
- **Friday 20st October:** Symposium at Warwick (tentative schedule, TBC)
11am-12.30pm: presentations
12.30pm: lunch
individual feedback over lunch
2-4.40pm: double seminar: TBC + TBC
4.40pm: wine
6pm: bus to Oxford

Objectives

After the course the students will have acquired the following skills:

- Familiarity with R and ability to implement mainstream computational algorithms
- Experience of reading and discussing code; understanding the benefits of producing clear well-written code
- Collaborating in understanding and implementing a research problem
- Critically assessing someone else's code and work, giving constructive feedback
- Receiving critical feedback and using it constructively to improve work
- Exposure to several mainstream topics in computational statistics
- Thorough familiarity with two of these topics (project done + project reviewed)
- Ability to present work and research

Assessment

Each team will produce a report of maximum 8 pages, and return code and graphs in the format of an R vignette. The report will be marked and there will be feedback on the code and presentation.

List of Papers

In 2017 the following papers will be presented along with projects. The list is subject to change, in particular additions, and will be confirmed when module starts.

1. (2017R) Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R. Least angle regression. Annals of Statistics, Institute of Mathematical Statistics, 2004, 32, 407-499
2. (2017R) Lauritzen, S.L. and Spiegelhalter, D.J. Local computations with probabilities on graphical structures and their application to expert systems, Journal of the Royal Statistical Society. Series B. 50(2): 157–224. 1988.

3. (2017K) J. Bierkens, P. Fearnhead, G. Roberts. The Zig-Zag Process and Super-Efficient Sampling for Bayesian Analysis of Big Data. preprint arXiv:1607.03188

4. (2017K) M. Vihola, J. Helske, J. Franks. Importance sampling type correction of Markov chain Monte Carlo and exact approximations. preprint arXiv:1609.02541

5. (2017K) D. Maclaurin, R.P. Adams. Firefly Monte Carlo: Exact MCMC with Subsets of Data. IJCAI 2015

6. (2017K) Iain Murray, Ryan Prescott Adams, David JC MacKay. Elliptical slice sampling. JMLR 2010

7. (2017R) Piotr Zwiernik, Caroline Uhler, Donald Richards. Maximum likelihood estimation for linear Gaussian covariance models. *JRSS-B*, 2016.