Course Outline

Summer Reading

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

`http://www.stats.ox.ac.uk/~evans/teaching.htm#R`

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 8th October:**
  - 10:00: introductions
  - 10:30: talk by Tom Nichols.
  - 11:30: break
  - 11:50: Computational problems in statistics - an introduction; Review of R; including calling C code, debugging; (Robin).
  - 14:00: Introduction to Monte Carlo methods (Krys).
- **Tue:**
  - 9:30 Computational problems in statistics (Robin, ctd).
  - 10:30 Presentation of four or five classic papers from computational statistics and related projects, about 20 minutes per topic. (Robin and George D).
  - 11:30: break
  - 12:00 Introduction to Git and Github (Robin)
  - 14:00 Exercises
  - 16:00 Round up
- **Wed:**

  - 10:00 Further classic presentations (Krys).
  - 11:30 Breaking into pairs or threes; assigning projects to groups

- **rest of week 1:** work on the project: produce code and draft reports
- **Mon 15th October: peer review day:**
  9:30: 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);
  14:00: review second team's work, give feedback (round 1)
  16:00: review second team's work, give feedback (round 2)
- **Wed:**
  - 12:30 pizza lunch and talk (Paul Hodgson, GLA)
  - 17:00 hand in code (in the form of an R vignette) and reports for marking
- **Thu:** work on presentation
- **Friday 21st October:** Symposium at Warwick (tentative schedule, TBC)
  11:00: presentations
  12.30: lunch
  individual feedback over lunch
  14:00: double seminar: Martyn Plummer (Warwick) and Ricardo Silva (UCL)
  16:40: wine
  18:00: 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

A subset of papers from the following list will be presented along with projects.

1. G Deligiannidis, A Doucet, MK Pitt, The Correlated Pseudo-Marginal Method. *JRSS-B*, to appear.
2. PA Jenkins, D Spano. Exact simulation of the WrightFisher diffusion. *Annals of Applied Probability*, 2017.
3. M. Newman. Network structure from rich but noisy data, *arXiv:1703.07376*, 2017.
4. D. Rubin. The Bayesian bootstrap. *Annals of Statistics*, 1981, 9, 130-134.
5. R.D. Shah and P. Buhlmann. Goodnessoffit tests for high dimensional linear models, *JRSS-B*, 2018.
6. C Sherlock, AH Thiery, GO Roberts, JS Rosenthal. On the efficiency of pseudo-marginal random walk Metropolis algorithms, *Annals of Statistics*, 2015.
7. G. Zanella and G.O.Roberts. Scalable Importance Tempering and Bayesian Variable Selection. *arXiv:1805.00541*, 2018
8. C.-H. Zhang and S.S. Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models, *JRSS-B*, 2014.