Part 0: Introduction to PHYS5513 (Computational Statistics for Physics)

Aaron Robotham and Danail Obreschkow

Hello!

Welcome to Computational Statistics for Physics. This will be a challenging but rewarding course for most participants.

This unit provides key tools for the statistical and numerical analysis of complex problems in physics and astrophysics. It includes a solid introduction to the **R** statistical programming language and contains many practical examples from cosmology and astrophysics, as well as some examples from general physics and medicine.

Who are we?

- Professor Aaron Robotham: aaron.robotham@uwa.edu.au / KJM G44
- Professor Danail Obreschkow: danail.obreschkow@uwa.edu.au / KJM 149

We are joint course coordinators and lecturers. We are both full-time research faculty at ICRAR-UWA with a particular passion for statistics and the role of deep analysis in astronomy.

If you need to get in touch with us, email is the best option. Whilst we can arrange to meet with you if necessary, please do not assume you can come to our office without prior warning (we will likely be busy with other things at a random given moment). When contacting as, we are happy for you to use are first names (we are not too formal about titles etc).

Lectures and Notes

The unit is presented in 18 sections, grouped into six parts of three sections each. All lectures and course activity will be in G40 (the Computer Lab in KJM). Each section corresponds to one 2h-lecture. Each part is taught by one lecturer, in an alternating manner as per schedule on the next page. The lecture content is supported by detailed lecture notes supplied in PDF and Rmarkdown format (one set of files per section). These notes are comprehensive and often contain more material than discussed during the lecture. The idea is not for students to memorize the notes in detail, but to understand the key concepts and remember where to look up the details. We strongly recommend that students read the lecture notes before the respective lectures.

Please note, since lectures take place in the G40 Computer Lab they will not be recorded, and no special provision will be made for those not attending the lecture physically. That said, multiple people have successfully navigated the course fully remotely just based on the detailed notes we provide.

Assessments

The assessment is based on six equally-weighted assignments, corresponding to the six parts of the course. There is no final exam. The assignments are distributed at the beginning of the respective parts of the course. Questions on how to solve the assignment can be asked during the lectures. The solutions of each assignment will be briefly discussed in class, and basic solutions put on-line.

Following UWA's 150hr/unit rule, students are expected to allocate their time roughly as follows: **36h** in-class for lectures, **54h** on reading/processing lecture notes (=3hr/section), and **60h** on completing the assignments (=10h/assignment).

The six assignments are distributed as PDF files and, in some cases, as Rmarkdown files. For each assignment, the solutions must be provided in a single **Rmarkdown** document, along with the associated compiled **PDF** file. Both files must be uploaded to the relevant link on the LMS Grade Centre before the due date. Late submissions will have points deducted as per UWA standard rules, except in extraordinary circumstances. Any requests for special consideration must be submitted in advance of any deadline. Hard copies or scans will not be accepted.

As per the current penalty system at UWA, if assignments are submitted within 48 hours of the deadline no penalty marks will be removed from the grade, but they will be noted as being 'late'. On the third day of late submission (between 48-72 hours) the penalty will be 15% of the total marks available, and an additional 5% will be removed on each subsequent day (i.e. additional 24 hour period from the original deadline). If an assignment is submitted after the 7 day limit (168 hours from the original deadline) then it will receive no marks.

Academic Integrity

Academic integrity is crucial. You are welcome to discuss assignments with peers and use online resources, including LLMs, but you must not copy or paraphrase others' work or employ services to complete assignments. Directly copying answers and code from any source is prohibited, even if modified (e.g., changing variable names, slight restructuring, or different comments).

Course Outline

Note the lecturers for each *Part* of the course, and individual lectures, might change from the outline below.

Part 1: R Language [ASGR]

- R basics
- Interfacing with compiled languages, including Rcpp examples, e.g. Logistic Map, Mandelbrot Set, Game of Life
- I/O handling & big data (read/write binary, ascii, FITS, HDF5, FST, arrow, feather, png/jpeg/bmp)

Part 2: Numerical Analysis [DO]

- Random numbers, non-elementary functions, interpolation/extrapolation, smoothing, density estimation
- Numerical calculus: integration, differentiation, differential equations
- Numerical optimisation (without MCMC)

Part 3: Statistical Concepts [ASGR]

- Probability trees and first glimpse at Bayesian statistics
- Statistical ensembles, populations/samples, probability density functions
- Statistical tests and sampling/resampling techniques

Part 4: Bayesian Inference [DO]

- Bayes' theorem, maximum likelihood and maximum a posteriori estimation, model evidence
- Laplace approximation: parameter uncertainties, error propagation, Fisher information, Jeffreys priors
- Number density functions, mass functions

Part 5: Advanced Applied Inference [ASGR]

- MCMC (write basic Metropolis Hastings sampler, *LaplaceDemon*)
- Correlations, linear modelling (*hyperfit*)
- Galaxy modelling (*profit*)

Part 6: Temporal and Spatial Statistics [DO]

- Fourier representation: Fourier series, FT, DFT; extended examples (e.g. image compression)
- 2-point and higher order statistics and common estimators
- Spherical harmonics and selected applications, including CMB