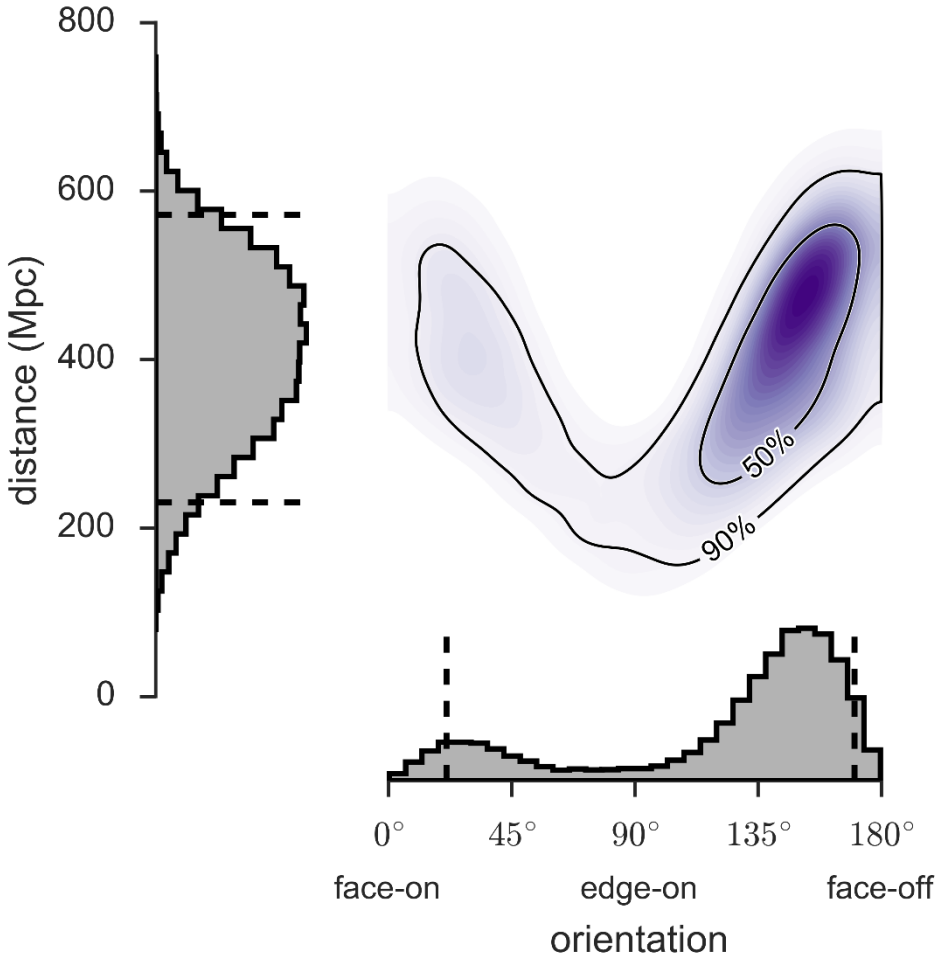
University of Glasgow | School of Physics & Astronomy



PHYS5001: Advanced Data Analysis

Course Information Guide

1 Course Details

Lecturers:	Prof Siong Heng	Schedule:	Self-study of lecture materials plus series of problem solving classes (typically 1 every 2 weeks)
SCQF Credits:	10	ECTS Credits:	5
Assessment:	Problems sheets (50%) Mock data challenge (50%)	Co-requisites:	None
Level:	Masters	Prerequisites:	None
Typically Offered:	Semester 2		

2 Course Aims

PHYS5001 Advanced Data Analysis is designed to give students a comprehensive introduction to the state-of-the-art in advanced data analysis methods as they are applied across a wide range of topics in physics and astronomy. Its aims are:

1. to equip students with a working knowledge of advanced data analysis methods to a level sufficient to permit their successful application to real data analysis problems, as might be encountered in students' own research projects;
2. to familiarise students with the key differences between a frequentist and Bayesian approach to data analysis: the assumptions upon which each approach is founded and the circumstances in which each is applicable;
3. to develop students' awareness of the current literature on advanced data analysis for the physical sciences, and the software available to support its application to real problems.

3 Intended Learning Outcomes

At the end of the course students should be able to:

1. Describe qualitatively the theoretical foundations of the nature of probability, in the context of both a frequentist framework (explaining clearly the meaning of this term) and a Bayesian framework (i.e. as a logical system for plausible reasoning).
2. Define what is meant by a probability density function (pdf), and cumulative distribution function (cdf), as well as various descriptive statistics (e.g. mean, median, mode, moments, variance, covariance) used to characterize pdfs and cdfs.
3. Apply the principles of least squares and maximum likelihood to formulate and solve simple line and curve fitting problems – using a matrix formulation where appropriate, and adapting the formulation to various cases and approximations (e.g. weighted least squares, correlated errors, non-linear problems).
4. Describe and apply the basic concepts of frequentist hypothesis testing, using the chi-squared goodness-of-fit test as an archetypal example.
5. Define in a Bayesian context the likelihood, prior and posterior distributions and their role in Bayesian inference and hypothesis testing, contrasting Bayesian and frequentist treatments of hypothesis testing.

6. Define the Fisher information matrix, its relation to the covariance matrix and its relevance to experimental design under the assumption of a Gaussian posterior.
7. Define the evidence and explain its role in Bayesian model selection, describing several numerical approximations to the evidence and their applicability.
8. Describe and apply efficient numerical techniques for generating random numbers and performing Monte Carlo simulations, including Markov Chain Monte Carlo methods.

4 Course Outline

Introduction and theoretical foundations; probability as a basis for plausible reasoning; conditional and marginal probabilities; Bayes' theorem; frequentist definition of probability.

Marginalisation; some important probability density functions (pdfs); measures and moments of a distribution; multivariate distributions; statistical independence; the bivariate normal pdf.

Frequentist parameter estimation; sampling distributions and estimators; the sample mean and sample correlation; the central limit theorem; the principle of least squares; ordinary and weighted least squares; extensions and generalisations.

The principle of maximum likelihood; least squares as maximum likelihood estimators; frequentist hypothesis testing; p-values, type I and type II errors; ROC curves and decision theory; Chi-squared goodness of fit tests.

Bayesian parameter estimation and hypothesis testing; sensitivity to prior information; Bayesian credible regions and frequentist confidence intervals; some worked examples.

Bayesian parameter estimation under the Gaussian approximation; the Fisher information matrix and covariance matrix.

Data compression and principal component analysis; defining probabilities; principle of insufficient reason and principle of indifference; Jeffreys' priors; maximum entropy and the definition of common pdfs.

Bayesian model selection; Occam factors and posterior odds ratio; worked examples; approximating the evidence.

Monte Carlo simulation methods; methods for generating and testing simple pseudo-random numbers; variable transformations; the probability integral transform; rejection sampling; genetic algorithms; Markov Chain methods; nested sampling and the evidence.

Fourier methods; Fourier transforms; convolution and correlation theorems; power spectral density; discrete Fourier transforms; fast Fourier transforms; the Nyquist-Shannon sampling theorem.

5 Further Information

Further information can be found on the course Moodle page and also using the links below:

- Course specification
- Reading list