| | |
|---|---|
| **Course Title** | ***Advanced methods in statistical data analysis*** |
| **Date** | ***October 19-23, 2009*** |
| **Registration deadline** | ***September 19, 2009*** |
| **Institute** | Niels Bohr Institute |
| **Guest teacher** | Dr. Wouter Verkerke (Nikhef , Amsterdam) |
| **Guest teacher** | Prof. Andrew Liddle (Sussex University, UK) |
| **Teachers from SCIENCE** | Dr. Steen H Hansen (NBI/DARK), Dr. Stefania Xella (NBI/HEP) |

# Course Plan:

## Description:

Data from experiments in high energy physics and observations in astrophysics demand nowadays a highly sophisticated statistical treatment. By inviting experts from both areas, we provide the students with the widest overview of the most advanced statistical methods. The course will cover the fundamental concepts of modern statistical data analysis, including examples derived from the two areas of science mentioned. The course is well suited and relevant for PhD students from a wide range of fields in Science, beyond the ones mentioned. The course will consist of lectures and practical problem-solving sessions (both calculations and computer exercises) and will last 5 full days.

For students showing active participation to the course, 5 ECTS points will be assigned. Registration can be done via email to xella@nbi.dk , hansen@dark-cosmology.dk. Deadline for registration is September 19, 2009.

## Content:

**Day 1-3**

lecturer: Dr. W.Verkerke

1) Basic Statistics

Mean, Variance, Standard Deviation. Gaussian Standard Deviation. Covariance, correlations. Basic distributions : Binomial, Poisson, Gaussian. Central Limit Theorem. Error propagation

2) Event classification

Comparing discriminating variables. Choosing the optimal cut. Working in more than one dimension. Approximating the optimal discriminant. Techniques: Principal component analysis, Fisher Discriminant, Neural Network, Boosted Decision Trees, Probability Density Estimate, Empirical Modeling

3) Estimation and fitting

Introduction to estimation. Properties of chi-2, Maximum Likelihood estimators. Measuring and interpreting Goodness-Of-Fit. Numerical stability issues in fitting. Mitigating fit stability problems. Bounding fit parameters. Fit validation studies. Maximum Likelihood bias issues at low statistics. Toy Monte Carlo techniques. Designing and understanding Joint fits. Designing and understanding Multi-dimensional fits.

4) Confidence interval, limits & significance

Probability, Bayes Theorem. Simple Bayesian methods and issues. Frequentist confidence intervals and issues. Classical hypothesis testing. Goodness-of-fit. Likelihood ratio intervals and issues. Nuisance parameters. Likelihood principle

5) Systematic uncertainties

Sources of systematic errors. Sanity checks versus systematic error studies. Common issues in systematic evaluations. Correlations between systematic uncertainties. Combining statistical and systematic error

and problem-solving (S.Xella assisting)

**Day 3-5**

lecturer: Prof. A.Liddle

1) Some cosmological background

This topic will discuss some of the current issues in cosmology demanding advanced statistical treatments, in order to provide focus and motivate some examples that will be used during the course. Note however that most of the remaining course material will be applicable to a wide range of scientific discplines.

2) Inference

A discussion of the underpinnings of statistical inference, particularly those of the Bayesian school.

3) Parameter estimation and Monte Carlo methods

Techniques for estimation of model parameters, likelihood analysis, Monte Carlo sampling methods, Metropolis-Hastings algorithm, machine learning.

4) Model selection and multi-model inference

Techniques for comparison of competing models, model simplicity and predictiveness, Bayesian model selection, computational approaches to model selection, inference in the presence of model uncertainty (multi-model inference), non-Bayesian methods and information theory.

5) Forecasting and experimental design

Quantifying experimental capability, optimizing experimental capability, parameter estimation and model selection approaches to optimization.

and problem-solving (S.Hansen assisting)

## Goal:

After the course, the students will have a detailed understanding of the fundamental concepts of modern statistical data analysis. They will also be able to use those concepts in solving concrete problems arising in data analysis, after having trained on some specific problems based on high energy physics or astrophysics data samples during the course.

## Pensum:

Some literature/material to be used in the course (additional will be provided before the

course starts):

Part I:

http://roofit.sourceforge.net/

https://twiki.cern.ch/twiki/bin/view/RooStats

Part II:

A.R. Liddle, ``Statistical methods for cosmological parameter selection and estimation," arXiv:0903.4210 [hep-th].

A.R.Liddle,``Information criteria for astrophysical model selection," Mon.Not.Roy.Astron.Soc.Lett. **377**(2007) L74, [arXiv:astro-ph/0701113].