Statistical Methods for Astronomers
Winter 2008
(12-02-2008 – 04-04-2007)

Lecturers: Russell Shipman (x7753) and Saleem Zaroubi
russ@sron.nl :ZG 276
saleem@astro.rug.nl :ZG 282

Course Times: Lecture: Tuesday: 11:15 - 12:45
Lecture: Friday: 11:15-12:45
Werkcollege: Wednesdays or Thursdays for an hour

(exam somewhere between 7th and 25th of April)

Course Place: ZG 161 for both lectures and exercises.

Resources:

- Practical Statistics for Astronomers, J.V. Wall and C.R. Jenkins (ISBN 0-521-45616-9)
- Statistics in Theory and Practice, Robert Lupton, (ISBN 0-691-07429-1)
- Numerical Recipes. Press, Teukolsky, Vetterling, Flannery (ISBN 0-521-43064-X)
- Kapteyn computing facilities.
- Saleem and Russ
-

Course description:
Each week we will have a lecture and work assignments. Some of the work assignments will require a bit of programming. There will also be a project due at the end of the course. This year the project will be a computer investigation which will culminate in a presentation and report. We'll discuss the project in more depth when the time comes.

Evaluation:

- Final Exam: 50%
- Project: 35%
- Work Assignments: 15% (includes class participation).

Syllabus for Statistical Methods for Astronomers

Week 1 (11-02-2008 – 15-02-2008 : Russ) Introduction:
At the end of the courses the students will be to:

1. Describe the purpose of statistics.
2. State the role of probabilities in decision making.
3. Describe the properties of a probability distribution function
4. Identify the Uniform, Normal, Poisson, Binomial and Cauchy p.d.fs
5. Identify a Bivariate Gaussian distribution.
6. Calculate the characteristic functions of these p.d.f.s
7. Define expectation values
8. Define standard measures for "center and width" of a distribution
9. Describe/calculate how p.d.f.'s change with various combinations of random variables
10. Calculate moments of a p.d.f using the characteristic function of that p.d.f.
11. Describe conditional probability.
12. Write Bayes Theorem
13. Properly formulate a problem using Bayes Theorem
14. Properly marginalize a variable.
15. Describe the Central Limit Theorem

Week 2 (18-02-2008-22-02-2008: Saleem) Monte Carlo
At the end of the courses the students will be to:

1. Describe situations where it is useful to "generate" your data.
2. Apply the bootstrap method to estimate uncertainties.
3. Create random numbers drawn from the uniform distribution.
4. Identify limits of computational methods (how many random numbers can be generated before repeating.
5. Create random numbers from a given p.d.f.
6. Identify Discrete Distributions
7. Marginalize a parameter from a distribution.
8. Simulate observed data
9. Describe the steps needed to simulate data

Week 3 (25-02-2008 to 29-02-2008: Russ) Statistics and Expectations:
At the end of the lectures the students will be to:

1. Define a statistic
10. Identify the differences between a statistic and a distribution parameter.
11. Identify the moments of a distribution with the Expectation values of the distribution.

12. Define the terms: *biased, consistent, closeness and robust* in terms of a statistic.
13. Define the covariance in terms of the Expectation value.
14. Describe the method for combining distributions. (including using the characteristic function.
15. Identify the distributions of some standard statistics: average, sample variance
16. Describe a correlation in terms of the Bivariate Normal distribution.
17. Describe a "Fishing Trip"
18. Describe ways in which a calculated correlation may be misinterpreted.
19. Write the estimator for this correlation coefficient and the distribution of this estimator.
20. Describe how Bayesian correlation testing differs from classical correlation testing
21. Describe a Non-parametric test for correlations
22.

Week 4 (03-03-2008 to 07-03-2008: Saleem) Numerical Fun with Correlations:
At the end of the courses the students will be to:

1. Define a partial correlation
23. Define Principal Component Analysis (PCA)
24. Apply PCA to a data set
25. Calculate Eigen values...
26. Describe how FFT's work.
27. Calculate a convolution of real data using an FFT
28. Calculate the power spectrum from an FFT

Week 5 (10-03-2008 to 14-03-2008: Russ) Hypothesis Testing

1. Describe the Basics of Hypothesis Testing
29. Apply t and F tests
30. Calculate the Bayesian Evidence to distinguish between two models
31. Identify and apply the following Non-parametric single sample tests
    1. Chi-square test
    2. Komogorow-Smirnov on-sample test
    3. one sample runs test of randomness
32. Identify and apply the following non-parametric two independent sampletests.
    1. Fisher exact test
    2. Chi-square two sample (k-sample-) test
    3. Wilcoxon -Mann-Whitney U test
    4. Kolmorogov-Smirnov two-sample test
33.

Week 6 (17-03-2008 to 21-03-2008 Russ and Saleem) Data Modeling

1. Define and apply the maximum likelihood method

34. Define and use the Hessian and relate this to the covariance matrix.
35. Describe the method of Lease Squares and identify how this is different than the MLE
36. Describe and apply a Bayesian likelihood analysis
37. Calculate the uncertainties in model parameters from the MLE, least squares and Bayesian analyses.
38. Determine the significance of model parameters
39. Describe and apply the minimum chi-square method.
40. Describe how to expand modeling using hierarchical models.
41. Identify how to calculate the evidence of a model and how to chose between two models.
42. Test a model fit to data

Week 7 (24-03-2008 to 28-03-2008: Russ and Saleem) The Project

1. Description/Development and Analysis

Week 8 (31-03-2008 to 04-04-2008: Russ and Saleem) The Project continued

1. More Analysis

43. Presentations
44. Hand in of project

Project includes the actual analysis, write up and presentation.

Exam: somewhere between 7th and 25th of April