ASTR 324: Introduction to AstroStatistics and Big Data in Astronomy

Željko Ivezić & Mario Jurić

University of Washington, Spring Quarter 2017

Location and Time: Tuesdays and Thursdays 2:30-3:50, **A210**

Office Hours: Any time when our office doors are open;
After Tue class, or Tue and Thu mornings are the best.

Grading: 10 homeworks, 8% each; final exam: 20%
key: >90%=A, >80%=B, >70%=C, >50%=D.

Class web site: https://github.com/uw-astr-324-s17/astr-324-s17

Reference textbook:
Ivezić, Connolly, VanderPlas & Gray: *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data* (Princeton University Press, 2014)
See http://press.princeton.edu/titles/10159.html

Learning Goals:
This course will introduce students to most common statistical and computer science methods used in astronomy and other physical sciences. It will combine theoretical background with examples of data analysis based on modern astronomical datasets. Practical data analysis will be done using python tools, with emphasis on astroML module (see www.astroML.org). While focused on astronomy, this course should be useful to all students interested in data analysis in physical sciences and engineering. The lectures will be aimed at undergraduate students and the main discussion topics will be based on Chapters 4 and 5, and selected topics from Chapters 6-10, from the reference textbook.

By taking this course, students will develop basic understanding of topics such as robust statistics, hypothesis testing, maximum likelihood analysis, Bayesian statistics, model parameter estimation, the goodness of fit and model selection, density estimation and clustering, unsupervised and supervised classification, dimensionality reduction, regression and time series analysis. Most of these topics will be applied in class homeworks to analysis of astronomical data.

Prerequisites: The students taking this class are required to have basic calculus and basic python skills, as well as basic scientific measurements and statistics skills at the level of a freshman lab.

Lecture format: New material will typically be covered during the first class in a week, while the second class in a week will be more focused on practical data analysis work.

Class Schedule (very tentative and subject to change)

- WEEK 1 (starting Mar 27): Introduction: syllabus, literature, python, matplotlib, astroML, scientific measurements.
- WEEK 2 (starting Apr 3): Introduction to statistics (probability, frequentist, distributions, CLT, robust statistics, hypothesis testing)
- WEEK 3 (starting Apr 10): Maximum likelihood and applications in astronomy.
- WEEK 4 (starting Apr 17): Bayesian statistics and introduction to MCMC
- WEEK 5 (starting Apr 24): Model parameter estimation and model selection
- WEEK 6 (starting May 1): Time series analysis
- WEEK 7 (starting May 8): Big data in astronomy
- WEEK 8 (starting May 15): Dimensionality reduction and regression
- WEEK 9 (starting May 22): Density estimation and clustering
- WEEK 10 (starting May 29): Unsupervised and supervised classification
- FINAL EXAM: June 6 (Tue): pizza and closed book exam

Homework

There will be ten homeworks, assigned on Thursdays, and due next Thursday. They will be centered on practical work using python and designed to test the weekly progress.

We will use modern software engineering tools, github and Jupyter notebooks for HW submission.