ASTR 598A: Introduction to Astrostatistics and Data-Intensive Astronomy

Mario Jurić

University of Washington, Autumn Quarter 2021

Location and Time: MW 2:30pm-3:50pm, Room B305 (and Zoom at `http://dirac.us/324`)

Office Hours: After Wednesday class

Grading: homeworks, 50%; final project: 40%; quizes: 10%.

Class materials: `https://github.com/uw-astrostats/astr-598a-au21`

Class JupyterHub: `https://dirac.us/hub598`

UW Astronomy Slack: `https://join.slack.com/t/uw-astronomy/signup`, then join `#astr-598-astrostats`

Textbook: Ivezić, Connolly, VanderPlas & Gray: *Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data*

Flipped classroom with online teaching:

This course will follow the flipped classroom model. In this method of teaching, you will listen to (prerecorded) lectures at home, and come to class (virtually, via Zoom) to engage in discussion, group work, and work on homeworks.

A typical cycle will run as follows:

- T+0 days: **by Friday 9am:** lectures for next week published on class YouTube channel. Please review them (as well as supplement them with chapters from the textbook) in time for a short quiz and survey on Monday. New homework will be assigned at the same time.
- T+3: **by Monday 12noon:** take the quiz, fill out the survey about what was not clear.
- T+3/5: **Monday and Wednesday class**: group discussion about outstanding questions, work on homeworks. Ideally, you will have finished your homework by the end of Wednesday class.
- T+7: **by Friday 9am:** lectures (and homeworks) for next week posted, and the cycle repeats...
- T+14: **by 9am the *following* Friday:** homework assigned two weeks ago are due.

Learning Goals:

This course will introduce you to concepts, tools, and and techniques from statistics and computer science that are essential for accurate and reproducible analysis of datasets, large and small. Through a series of lectures and hands-on problems, we will learn about elementary statistics, maximum likelihood methods, Bayesian probability and inference, MCMC methods, databases, and time series analysis. Practical data analysis will be done using Python, including astroML, astropy, astroquery and others.

The goal of this course is to give you the basic skills necessary to understand and correctly analyze rich datasets, from Kepler to LSST. It will also give you the theoretical prerequisites needed to successfully proceed to ASTR 597 Machine Learning in Astronomy (to be offered in the Winter quarter).

Prerequisites:

Students taking this class are required to have basic calculus and Python skills.

Topics:

- WEEK 1: Getting started with online/flipped classroom learning
- WEEK 2: Introduction to probability and statistics I
- WEEK 3: Introduction to statistics II
- WEEK 4: Maximum likelihood and applications in astronomy
- WEEK 5: Bayesian inference and model selection
- WEEK 6: Markov chain Monte Carlo methods
- WEEK 7: Project Proposals and Discussions
- WEEK 8: Dimensionality reduction
- WEEK 9: Time series analysis
- WEEK 10: Project Work I
- WEEK 11: Project Work II

Homeworks:

There will typically be a homework each week, assigned on Friday (9am), and due **two weeks later** on Friday (9am). The homeworks will focus on practical work using Python, designed to exercise what we've learned in the week after the homework is assigned. All homeworks will involved writing Jupyter notebooks.

Quizzes:

Multiple-choice quizzes will be every due every Monday at noon (12pm). They will cover key concepts discussed in lectures posted the previous Friday.

Final Projects:

You will propose a project (a piece of software, or an analysis) to build or improve using the techniques and libraries we'll learn about in the course. Ideally, this is something that helps with your research. Keep your eyes open for ideas!

Timeliness policy:

- All homeworks are due **two weeks** after being assigned.
- Homeworks turned in late (up to a week) will receive a 20% point deduction.
- **Homeworks turned in more than a week late receive a 50% point deduction.**.

Privacy/FERPA statement:

This course is scheduled to run synchronously at your scheduled class time both in-person and via Zoom. These Zoom class sessions **may** be recorded. The recording will capture the in-class audio, video and computer screen. Audio and video of remote attendees will be recorded if they share their computer audio and video during the recorded session. The recordings will only be accessible to students enrolled in the course to review materials. These recordings will not be shared with or accessible to the public.

The University and Zoom have FERPA-compliant agreements in place to protect the security and privacy of UW Zoom accounts. Students accessing remotely who do not wish to be recorded should: a) Change their Zoom screen name to hide any personal identifying information such as their name or UW Net ID, and b) Not share their computer audio or video during their Zoom sessions.