ISYA 2020 Cape Town

Course title: Astrostatistics Lab

Lecturer: Juan Rafael Martínez Galarza, Center for Astrophysics | Harvard & Smithsonian

Description: Astronomers are currently gathering data at higher rates than ever before, and this rapid increase in astronomical data volume will likely continue over the next decade or two . As an example, upcoming facilities such as the Large Synoptic Survey Telescope (LSST) and the Square Kilometer array will gather data at rates of several terabytes per second. The era of big data in astronomy (but also proper data analysis and model selection in the small data regime) requires sophisticated statistical inference methods, and the astronomer of today requires a solid knowledge of statistics and its application to astronomical measurements, large datasets, and discovery of the unexpected. In this practical lecture series I will provide an overview on basic astrostatistics concepts, and will focus on practical applications and hand-on activities. The lectures will cover topics from very basic questions about how to interpret photometric measurement errors, to more advanced topics on how we can use statistics to build machine learning models.

SYLLABUS

Lecture 1. Uncertainties and errors in astronomy

- The basics of probability
- Probability rules
- Random variables, expectation values and covariances
- Distributions and samples. The stellar Initial Mass Function as a probability distribution.
- Lab: Counting photons in X-ray astronomy: Poisson noise

Lecture 2. Fitting models to data

- Test of hypotheses (F-test, etc.)
- What is a model? Measuring the universe with photometric redshifts.
- Fitting statistics (chi squared, cash statistics, etc.)
- Maximum likelihood estimate
- p-Values and confidence intervals
- Lab: Fitting X-ray images with Sherpa.

Lecture 3. The world of Bayes

- Frequentism vs. Bayesianism
- Priors, posteriors, and posterior predictives.
- Bayesian modeling: probabilistic astronomical catalogs.
- Bayesian solution to counting X-ray photons.
- Markov-Chain Monte Carlo.
- Lab: Bayesian fits to X-ray spectra.
-

Lecture 4. Time domain and X-ray astronomy

- Introduction to time domain astronomy
- Time series
- Time domain surveys
- Variability in X-ray data.

- Period estimation for variable sources
- Feature extraction
- Lab: Building automatic classifiers for variable sources.

Lecture 5. Introduction to Machine Learning

- The statistical foundations of machine learning
- The logistic regression as a classifier
- Introduction to rural networks.
- Lab: convolutional neural networks for morphological classification of galaxies,

Requirements: Projector for the lectures, a computer lab hopefully with one computer per student (or at least one computer per every two students). Software requirements: Anaconda python, Chandra X-ray data analysis tools (CIAO and Sherpa).