Maestría en astrofísica
IRyA-UNAM

Temas Selectos de Astronomía: estadística

Omaira González Martín, responsable de posgrado jposg@irya.unam.mx
Karin Hollenberg, administración de posgrado *k.hollenberg@irya.unam.mx*
Página de posgrado: https://posgrado.irya.unam.mx
Contacto para pedir información: posgrado@irya.unam.mx

Temas Selectos de Astronomía: estadística

Sundar Srinivasan

The tremendous volume of data delivered by ground- and space-based telescopes, coupled with the ever-increasing computing power, has created a problem: most data will never be seen by a human, either due to lack of human analysts or due to automated processing techniques. Students and researchers have been slow to catch up with the advances in statistical and programming techniques required to handle such large datasets, partly due to the fact that such training is not available as part of graduate programs in astronomy and astrophysics. This course aims to bridge this gap.

The student will gain experience in classical statistical methods and as well as in current topics such as Bayesian techniques. The course will focus on the programming-based analysis of data using Python, and will also discuss standard topics such as fitting models to real astronomical data.

Thematic Content

1. **Basic probability theory and statistics**.- Conditional probability, and Bayes' theorem. Bayesian inference versus frequentist inference.

2. **Random variables and probability distributions**
   - Populations vs. samples. Discrete and continuous probability distributions and their moments. The Central Limit Theorem.
   - The cumulative distribution function. Expectation values, variance, standard deviation, covariance, correlation coefficients, correlated and uncorrelated random variables.
   - Distributions of functions of random variables.

3. **Statistical inference**
   - Estimators: mean, variance, bias. The bias-variance tradeoff
   - Statistical models. Likelihood: frequentist and Bayesian interpretations. The Maximum Likelihood Estimator, Fisher Information, and the Cramér-Rao Theorem. The \chi^2 and \chi distributions.
   - The Student's t distribution and the t statistic.
   - The Bootstrap Method.
   - Confidence sets and (one- and two-sided) intervals. Confidence bands.
   - Hypothesis testing. The p value. Parametric and nonparametric tests. The empirical distribution function. The KS test

Omaira González Martín, responsable de posgrado jposg@irya.unam.mx
Karin Hollenberg, administración de posgrado *k.hollenberg@irya.unam.mx*
Página de posgrado: https://posgrado.irya.unam.mx
Contacto para pedir información: posgrado@irya.unam.mx

4. **Visualising data**
    - Five-number summaries and the box-and-whisker plot. Histograms. Nonparametric density estimation and Kernel Density Estimates.

5. **Bayesian inference**
    - Priors and prior selection. Bayesian point/location and interval estimates. The Highest Posterior Density Interval. Prior- and data-dominated posteriors. The Jeffreys Prior.
    - Model selection, complexity, and information criteria. The posterior predictive distribution Sampling techniques
    - Monte Carlo. Rejection and importance sampling. Markov Chains and Markov Chain Monte Carlo. The Metropolis-Hastings Algorithm. Implementations in astronomy: emcee.

6. **Regression**
    - Heteroskedastic uncertainties. Ordinary least-squares. Detection limits. Outliers and robust regression. Bayesian outlier rejection (marginalization). Goodness-of-fit. Bivariate measurement uncertainties and intrinsic scatter.

For the advanced topics such as MCMC and Bayesian modeling, the plan is to have a couple of experts as guest speakers.