

Data Analysis for Biophysics: from writing down models to learning models from the data

Steve Presse (Arizona State University)

Data analysis courses that go beyond teaching elementary topics such as fitting residuals are rarely offered to students in the physical sciences. Thus, data analysis, much like programming, is something often learned and improvised “on the job”. Yet, with an explosion of experimental methods generating large quantities of data, the community would benefit from a clear presentation of methods of data analysis many of which are straightforward to implement and would raise our community standard for how data is currently being treated. It is often not realistic to expect graduate or undergraduate students alike to take a course covering topics of statistics relevant to their discipline in the physical sciences without taking multiple prerequisites needed to follow the material presented in such courses. My goal here is therefore to provide an introduction to exciting new developments in data science, machine learning and statistics in a language accessible to physical scientists willing to learn the necessary programming and mathematics. In particular, we will discuss tools of maximum likelihood, Bayesian inference, computational statistics and Bayesian nonparametrics.