Inference — A Python Package for Astrostatistics

Tom Loredo (Dept. of Astronomy, Cornell University), Alanna Connors (Eureka Scientific), and Travis Oliphant (Dept. of Electrical & Computer Engineering, Dwight B. Young University)

Project Overview

Motivation

Many advanced astrostatistics methods are conceptually simple despite being computationally complex. Competing methods of very different levels of sophistication are often similar from an end-user’s perspective. The principle obstacle to the use and understanding of advanced methods is the art of statistical computing—the computational tricks needed to implement advanced methods.

Goal: Eliminate this obstacle!

Example—Fitting binned spectral data from data contaminated with measured background:

- Estimate χ using background-subtracted data
- Maximize a Poisson counting process likelihood marginalized over a bin-by-bin background model

These are quite similar from a user’s perspective. One must (1) Define a parameterized signal model; and (2) Optimize a scalar function of the model’s parameters. Analysts should not be prevented from trying the (more exact) likelihood approach simply because efficient computation of the likelihood requires unconventional computational “tricks.”

Main Features

The Inference project is making advanced astrostatistics methods accessible to astronomers through the following project components:

- The Inference package—Two software components
  - Library: A deep and broad collection of self-contained functions and objects implementing methods tailored to astronomers’ needs. Where possible, it includes multiple methods in each problem class, e.g., frequentist/Bayesian
  - Parametric Inference Engine: A framework for analyzing parametric models allowing use of multiple methods (χ², likelihood, Bayes) with a unified interface
- Use of a modern “very high level” (VHL) computer language: Python
- Single implementation facilitates upgrades/development (e.g., spreading resources across implementations in several languages)
- Python’s VHL features speed development, facilitate testing

- Parametric Inference Engine
  - Optimized simplicity allows easy access both to new users and to astronomers using PyRAF
  - Outreach
    - This project organizes and sponsors astrostatistics speakers and seminars at astronomy and astrophysics conferences (like HEAP)
    - Selected methods described in project-sponsored talk will be included in the Python package

Language Characteristics

- A general purpose language with a rich standard library
- Very simple syntax—resembles “pseudo code”
- Use interactively, or via script/modules
- Object oriented, with a very simple object model—facilitates high level interfaces, modularity
- Practical rather than ‘pure’—Selected capabilities of various paradigms (e.g., functional programming, list comprehensions, metaclasses)
- Sophisticated and fast scientific computing capability
- Easily extendible/extendable with C++/Fortran
- Open source, cross-platform, active & growing user community
- Named for the British comedy show, not the snake!

Scientific Computing With Python

- Array computing—From numerical linear algebra to 3-d visualization
- Many advanced astrostatistics methods are included
  - Robust estimation/outlier detection (M-estimators, Bayesian outlier detection)
  - Detection/measurement of periodic signals
  - Nonparametric time series analysis (QPOs, long-memory processes)
  - Linear & nonlinear regression
  - Fitting with correlated errors
  - Detection/measurement of periodic signals
  - Robust estimation/outlier detection (M-estimators, Bayesian outlier detection)

Simple Example

Rayleigh statistic for period searching in arrival time data:

```
InferenceExample.py

InferenceExample()
```