2000 years of astrostatistics

Ancient Greeks through the 19th century:

The astronomers *were* the statisticians: Hipparchus, Brahe, Legendre/Laplace/Gauss, Kapteyn/Newcomb/Airy

Late-19th to late-20th century:

The lost century … statistics moves towards human affairs and astronomy towards astrophysics. Astronomers missed the statistical revolution, statisticians missed cosmology.

1970-80s:

Isolated progress … Lynden-Bell-Woodroofe estimator for truncated data, Stellingwerf & Lomb-Scargle periodograms for irregularly spaced data, survival analysis, galaxy clustering, etc.

1995-present:

Resurgence of a vanguard of astrostatistics: Bayesian inference, Poisson processes, data mining, model selection, compressive sensing, etc.

SCMA V panel discussion Eric Feigelson

Yet the ***average*** astronomical paper today uses a narrow suite of older methods, often poorly (minimum-χ^2 regression, Kolmogorov-Smirnov test, Likelihood Ratio Test, single-linkage hierarchical clustering, etc.)

Usage of classification and clustering methods

Method	Google[1]	ADS[2]
Data mining	16.6 million	20
Neural network	7.1 million	135
Machine learning	5.6 million	20
Hierarchical clustering[3]	950 thousand	400
Support Vector Machine	870 thousand	15
Linear discriminant analysis	500 thousand	5
Random Forest	260 thousand	3
Bayesian classifier	230 thousand	15
Classification and Regression Tree	160 thousand	0
k-nearest neighbor	110 thousand	9
Model-based clustering	56 thousand	2

A blatant advertisement:

Modern Statistical Methods for Astronomy with R Applications

E.D. Feigelson & G.J. Babu Cambridge Univ Press late-2011

Astrostatistics: An astronomer's view

The application of statistics to scientific data is not a straightforward, mechanical enterprise. It requires careful statement of the problem, model formulation, choice of statistical method(s), calculation of statistical quantities, and validation of results. This is a messier, but much more interesting, process than conducted by most astronomers.

Astrostatistics should be a significant niche cross-disciplinary field, and a few percent of astronomers should be specialist astrostatisticians (cf. astrochemists, instrumentalists).

Modern statistics is vast in its scope and methodology. This makes it difficult to find what may be useful, but gives enormous capabilities. Some procedures are based on mathematical proofs while others are not; astronomers should know the difference.

Interpreting a statistical result is not always obvious. We are scientists first! Statistics is only a tool. We should be knowledgeable in our use of statistics and judicious in its interpretation.

A vision of astrostatistics in 2025 …

The undergraduate/graduate curriculum for astronomy includes a full year of statistical inference and methodology oriented towards physical science. Dozens of young astronomers have M.S. degrees in statistics and computer science. Science based on petabyte/exabyte datasets use modern methods effectively and thoughtfully. Astronomical papers reference statistics monographs.

Methodologies for problems that seemed challenging in 2011 -- multivariate heteroscedastic measurement errors, irregularly-spaced time series, faint source detection, time series classification -- are now well-established. Astronomers regularly use hundreds of methods coded in **P**, the successor to **Q** and **R**.

Thirty well-funded cross-disciplinary research groups in astrostatistics and astroinformatics on three continents push frontiers of astrostatistical methodology. ***Statistical Challenges In Modern Astronomy*** meetings are held annually with hundreds of participants, alternating between Penn State and locations with lovely beaches.