Astronomical Data Analysis Software and Systems IV
ASP Conference Series, Vol. 77, 1995
R. A. Shaw, H. E. Payne, and J. J. E. Hayes, eds.

Statistical Consulting Center for Astronomy

E. D. Feigelson

Department of Astronomy & Astrophysics, Penn State University, University Park PA 16802

M. G. Akritas, J. L. Rosenberger

Department of Statistics, Penn State University, University Park PA 16802

Abstract. We announce the formation of a Statistical Consulting Center for Astronomy (SCCA), designed to provide prompt high-quality advice on statistical and methodological issues to the astronomical community. Questions should be sent by e-mail to *scca@stat.psu.edu*. Questions and answers can be examined on the World Wide Web[1].

1. Introduction

The astronomer extracting scientifically useful information from astronomical data often encounters complex and subtle problems. Statistical techniques such as least-squares model fitting, Kolmogorov-Smirnov two-sample test and χ^2 goodness-of-fit test can be applied to many simple situations, but are inadequate for other problems. A few examples of such data analysis problems are: satellite surveys with flux limits and nondetections; discrimination between stars and galaxies in digitized optical surveys; detection of weak sources in photon-counting detectors with variable backgrounds; characterization of quasi-periodic or stochastically variable objects; identification of filaments and voids in anisotropically clustered galaxies; analysis of the Lyman-α forest in quasar spectra; repeated application of calibration regressions in the cosmic distance scale; and error analysis in all of these situations.

The field of mathematical statistics and its many areas of application (biometrics, econometrics, chemometrics, geostatistics, quality control, etc.) have made huge advances in recent decades. Mathematics libraries have dozens of journals and hundreds of monographs on specialized problems in statistics that are rarely if ever read by the astronomer. The problem encountered by an astronomer has often been addressed, and perhaps clearly resolved, by statisticians working in other fields. In other cases, the astronomical problem is methodologically unique, and its treatment might challenge a top statistician specializing in the relevant field.

[1]http://www.stat.psu.edu/scca/homepage.html

We have created the Statistical Consulting Center for Astronomy (SCCA) to help bridge the wide gap between the astronomical and statistical communities. Through the SCCA, astronomers can ask a team of statisticians questions about the data analysis problems they are facing today. If a good solution is readily known, the SCCA will respond rapidly with an answer and guidance into the appropriate statistical literature. If the problem is particularly tricky or important, the SCCA will seek out top quality statisticians to consult with, and possibly collaborate with, the astronomer.

The need for improved statistical treatment of astronomical data is clear. A scan of the *Astronomy & Astrophysics Abstracts* indicates that 100–200 papers are published annually are principally concerned with methodological issues in the astronomical literature, and dozens of additional observational papers have discussions of statistical issues. Statistical issues arising in astronomical data analysis have been presented at a growing number of conferences (e.g., Jaschek & Murtagh 1990; Feigelson & Babu 1992; Subbarao 1995; various ADASS and European workshop proceedings). Yet except for the 1991 Penn State conference, there has been little involvement of the academic and professional statistical community in addressing the problems arising in astronomy.

2. Operation of the SCCA

The SCCA is a team of Penn State faculty with interest and expertise in statistical problems arising in astronomical research. The Center has contacts with experts in the international statistical community. The goals of the Center are to: (1) address the immediate statistical needs of astronomers by providing prompt high-quality statistical advice, (2) make publicly available questions and answers for the benefit of the wider astronomical community, and (3) encourage interdisciplinary collaboration between the fields of statistics and astronomy.

Any individual in the astronomical community can submit a question to the SCCA: a graduate student preparing a dissertation; a scientist confronted with a tricky data set, preparing or revising a paper for publication; a scientist preparing software for an instrument or a data analysis software system; or a scientist organizing a major observational program. Incoming questions are reviewed by the members of the team and colleagues in the Department of Statistics at Penn State. Many questions will be answered in-house, but particularly complex and important problems will be sent to top-ranked experts worldwide. The turn-around time for answering straightforward problems should be no more than three weeks. Summaries of questions and answers will be made publicly available through the Internet/WWW and publications.

The operation of the SCCA is partially supported by the NASA Astrophysics Data Program starting in fall 1994. Initially, consulting can be free of charge to U.S. astronomers. However, we strongly encourage questioners to pay a nominal fee for the service. This will ensure the continuance of the Center into the future, and the availability of top-quality external consultants.

When a question for the SCCA arises, astronomers should send e-mail to *scca@stat.psu.edu* or FAX the Center at (814) 863–7114. Questions and answers will be available by anonymous ftp at *ftp.stat.psu.edu* (cd to the *pub/scca* directory) and on the World Wide Web at the SCCA Homepage.

3. Some Early Questions & Answers

Q: Can a partial correlation coefficient be applied to data with upper limits?
A: One can construct a partial correlation coefficient for censored data using (say) the generalized Kendall's τ bivariate coefficient implemented in the ASURV package (LaValley et al. 1992), but no tests of significance are available. In fact, no significance testing method is available for the partial Kendall's τ even with uncensored data (Hettmansperger 1984, p. 208). For uncensored data, we recommend instead either multiple regression (Murtagh & Heck 1987) or Pearson's linear partial correlation coefficient (Anderson 1984). Unfortunately, the extension of multivariate analysis to censored data has proved to be quite difficult and there are no available methods. Thus, no fully satisfactory answer to your question exists, but an expert in the field has promised to work on developing a method for testing the hypothesis that the partial Kendall's τ is zero.

Q: How can one assess the likelihood and amplitude of variability of an X-ray source from *ROSAT* observations consisting of 20 disjoint good time intervals? 100-1000 total counts are collected, which is a bit low for the χ^2 test.
A: If you can confidently assume that the underlying distribution of counts follows a Poisson distribution, we recommend the likelihood ratio test. You want to test that $\lambda_1 = \cdots = \lambda_{20} = \lambda$, where λ_i times the exposure time t_i gives the expected counts $E(X_i)$ in the i-th interval. The likelihood ratio statistic is

$$L_R = 2 \sum_{i=1}^{n} X_i log(\frac{X_i}{\hat{\lambda} t_i}). \tag{1}$$

If the hypothesis of no variability is true, then L_R has a χ^2 distribution with 19 degrees of freedom. Thus, the null hypothesis is rejected at significance level $\alpha = 0.01$ if $L_R > 36.2$. If the hypothesis of constancy is rejected, the amplitude of variability can be examined from the estimated parameters $\hat{\lambda}_i = X_i/t_i$. The likelihood ratio test is presented in Hogg & Tanis (1993), and its use in astronomy under the Poisson hypothesis is discussed by Cash (1979).

Q: Consider two clusters of galaxies, one with $N_1 = 80$ galaxies with 40% spirals and the other with $N_2 = 120$ galaxies with 70% spirals. Is the spiral fraction difference significant?
A: Let $\hat{p}_1 = X_1/N_1$, and $\hat{p}_2 = X_2/N_2$ be the two proportions. We can suggest two test statistics for determining if the proportions are significantly different (Arnold 1990; Miller, Freund, & Johnson 1990):

$$T_1 = \frac{\hat{p}_1 - \hat{p}_2}{[\frac{\hat{p}_1(1-\hat{p}_1)}{N_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{N_2}]^{1/2}} \tag{2}$$

$$T_2 = \frac{\hat{p}_1 - \hat{p}_2}{[\hat{p}(1-\hat{p})(N_1^{-1} + N_2^{-1})]^{1/2}} \tag{3}$$

where $\hat{p} = (X_1 + X_2)/(N_1 + N_2)$. Under the null hypothesis, both statistics have a normal distribution with mean zero and variance one. T_2 is equivalent to Pearson's χ^2 and is more commonly used, though its applicability is limited to testing the null hypothesis. T_1 is a Wald-type statistic and can be used to give confidence intervals for the true difference $p_1 - p_2$. For the problem at hand, the two proportions would be declared significantly different at significance level $\alpha = 0.01$ if $|T_2| > 2.58$. Miller, Freund, & Johnson (1990) also consider a k-sample version of this statistic.

Q: I am teaching a graduate course on astronomical techniques, and would like to include a short section on Bayesian analysis. Can you suggest a general reference?
A: An excellent review of Bayesian inference in astronomy is given by Loredo (1992) and further applications are discussed in Ripley (1992). Background references might include Lindley (1965) and Howson & Urbach (1993).

Acknowledgments. The SCCA is partially funded by NASA grant NAS5-32669.

References

Anderson, T. 1984, An Introduction to Multivariate Statistical Analysis (New York, Wiley)

Arnold, S. 1990, Mathematical Statistics (Englewood Cliffs, Prentice-Hall), p. 386

Cash, W. 1979, ApJ, 228, 939

Feigelson, E. D., & Babu, G. J., eds. 1992, Statistical Challenges in Modern Astronomy (New York, Springer-Verlag)

Hettmansperger 1984, Statistical Inference Based on Ranks (New York, Wiley)

Hogg, R., & Tanis, E. 1993, Probability and Statistical Inference (Macmillan)

Howson, C., & Urbach, P. 1993, Scientific Reasoning: The Bayesian Approach (Chicago, Open Court)

Jaschek, C., & Murtagh, F. (eds.) 1990, Errors, Bias and Uncertainties in Astronomy (Cambridge, Cambridge Univ. Press)

LaValley, M., Isobe, T., & Feigelson, E. 1992, BAAS (Software Report), 24, 839

Lindley, D. 1965, Introduction to Probability and Statistics from a Bayesian Viewpoint, 2 vols., (Cambridge, Cambridge Univ. Press)

Loredo, T. 1992, in Statistical Challenges in Modern Astronomy, eds. E. D. Feigelson & G. J. Babu (New York, Springer-Verlag), 275

Miller, I., Freund, J., & Johnson, R. 1990, Probability and Statistics for Engineers (Englewood Cliffs, Prentice-Hall), p. 282

Murtagh, F., & Heck, A. 1987, Multivariate Data Analysis (Dordrecht, Kluwer)

Ripley, B. D. 1992, in Statistical Challenges in Modern Astronomy, eds. E. D. Feigelson & G. J. Babu (New York, Springer-Verlag), p. 329

Subbarao, T., ed. 1995, Applications of Time Series Analysis to Astronomy and Meteorology (New York, Chapman-Hall)