

# IMPROVING THE STATISTICAL METHODOLOGY IN ASTRONOMY

E. D. FEIGELSON

*Department of Astronomy & Astrophysics, Pennsylvania State University, University Park PA 16802, USA*

G. J. BABU

*Department of Statistics, Pennsylvania State University University Park PA 16802, USA*

The fields of astronomy and mathematical statistics were once intimately intertwined, later very distant, and now approaching each other with new needs and perspectives. Innovative approaches to astronomical data analysis problems are illustrated with parameter fitting in X-ray spectroscopy. Chi-squared minimization encounters several problems, and alternatives like nonparametric EDF tests, maximum likelihood estimation, finite mixture models and bootstrap resampling are suggested. Increased collaboration and consultation with statisticians is recommended for astronomical projects encountering difficult data analysis challenges.

## 1 The Reemergence of Astrostatistics

Few modern astrophysicists are aware that astronomers played a seminal role in establishing the foundations of mathematical statistics. For centuries, cosmologists were confronted with the task of combining discrepant data to constrain parameters of nonlinear models. Ptolemy apparently minimized the maximum discrepancy between data and model. In the Middle Ages, there was concern that averaging measurements would increase, rather than decrease, the accuracy of the result. Tycho Brahe conducted systematic studies showing that averaging measurements reduces the errors (Sheynin 1973).

Galileo chastised Chiaramonti for ignoring measurement errors and deleting undesired data concerning the parallax of the supernova of 1572. This was of great importance, as Chiaramonti concluded that the 'new star' lay closer than the Moon while Galileo showed it lay far beyond the planetary spheres. He advocated minimizing the sum of the absolute deviations of the observations from the model; today we would call this a form of robust estimation. After Newton, Legendre developed the least squares method and Gauss defined the normal error distribution in the context of celestial mechanics studies. Huygens, Halley, Lambert, Young, Quetelet and Bessel all made important contributions to both astronomy and statistics (Stigler 1986; Hald 1990).

Strong links between the two fields faded in the late-19th and 20th centuries. Astronomers turned towards the new physics – electromagnetism, fluid mechanics and especially quantum mechanics – for deep insights into celestial phenomena. Statisticians found challenges in social sciences like demography, psychology, and economics, as well as applied fields such as agriculture, medicine and manufacturing. Two isolated achievements shine during this period: the monograph *Statistical Astronomy* by Trumpler and Weaver (1953), and a series of *Astrophysical Journal* papers on galaxy clustering by the distinguished statisticians Jerzy Neyman and Elizabeth Scott.

However, the past decade has witnessed a resurgence of interest in statistical issues among astronomers. Conferences devoted to astrostatistics have been held in Strasbourg (Rolfe 1983; Jaschek & Murtagh 1990), and cross-disciplinary conferences entitled *Statistical Challenges in Modern Astronomy* at Penn State (Feigelson & Babu 1992; Babu & Feigelson 1997). The subject has been addressed in these Erice *Data Analysis in Astronomy* and the U.S. *Astronomical Data Analysis Software and Systems* conferences. We have recently produced a brief monograph *Astrostatistics* designed to introduce scholars in astronomy and statistics to the other field (Babu & Feigelson 1996).

The potential applications of advanced statistical methods in astronomy is vast. In general, astronomers will find that statistics often provides several ways of approaching a given problem, each with its own advantages and limitations. Thus, which the mathematics behind a statistical procedure may be rigorous, its applications to a specific problem does not give immutable results. For example, the *ASURV* statistical package (Feigelson & Nelson 1985; LaValley et al. 1992) provides six different tests giving the probability that two univariate samples with censoring (i.e. upper limits due to nondetections at the telescope) are drawn from the same parent distribution. The procedures were developed for actuarial, biometrical and industrial applications and have unfamiliar names (the Gehan, logrank, Peto-Peto and Peto-Prentice tests). Each test assumes a different but reasonable weighting for the nondetections.

## 2 An illustration: Parameter estimation for X-ray spectroscopy

X-ray astronomy is entering its fourth decade, and soft X-ray spectroscopy is becoming an increasingly important element of high energy astrophysics. Most celestial X-ray sources have plasma at temperatures of  $10^6 - 10^7$  K which emit both a bremsstrahlung continuum and numerous spectral lines from many L- and K-shell electron transitions of carbon through iron. In many cases, the plasma is not isothermal; either a range of temperatures (e.g. stellar coronae) or non-equilibrium conditions (e.g. supernova remnants) are present. In

addition, many sources exhibit absorption lines and edges due either to *in situ* cold material (e.g. molecular torus around a Seyfert galaxy nucleus) or Galactic interstellar absorption. The astrophysical models for X-ray spectra can thus be very nonlinear and complex with continuum, emission line and absorption line/edge components. The scientific goals are to determine what components are present and to characterize quantitatively the physical properties (temperatures, densities, elemental abundances, ionization states) of the emitting and absorbing regions.

The quality of spectral data has historically been poor, but is rapidly improving. The dominant detector during the 1970-80s (e.g. *Einstein* IPC, *ROSAT* PSPC and *ASCA* GIS) were proportional counters with energy resolution around  $E/\Delta E \simeq 3 - 5$ . Information on the complex astrophysical spectra was reduced to two or a few hardness ratios. The *ASCA* SIS and forthcoming *AXAF* ACIS detectors are CCDs with intermediate resolution around  $E/\Delta E \simeq 20 - 50$ . Here strong individual emission lines can be resolved, but a great deal of line blending is present. The *Einstein* FPCS, *AXAF* transmission gratings and the X-ray bolometers planned for *Astro - E* will have  $E/\Delta E \simeq 100 - 1000$ . Here, hundreds of individual electron transitions should be resolvable and, providing sufficient photons can be acquired to fill the high resolution spectrum, the scientific goals should be directly achievable.

### 2.1 Chi-squared minimization and its deficiencies

The standard procedure for estimating astrophysical parameters from all of these spectra is to bin the data so each of  $i = 1, \dots, k$  energy channels has  $N_i > 5$  photons, assume a model  $M(\vec{\theta})$  where  $\vec{\theta}$  is the vector of astrophysical parameters, and calculate the parameters that minimize the weighted sum of squares defined by

$$\chi^2_{\nu}(\vec{\theta}) = \left( \sum_{i=1}^k [N_i - M_i(\vec{\theta})]^2 / \sigma_i^2 \right) / \nu. \quad (1)$$

Here  $\sigma_i$  is the Gaussian or Poisson error of the counts and  $\nu$  are the degrees of freedom.  $\chi^2(\vec{\theta})$  is calculated for a grid of models and the model giving  $\chi^2_{min}$  is considered the best fit. Parameter confidence bands are computed using  $\chi^2_{min} + const$  contours, where the constants depend on the number of 'interesting' parameters were derived by Lampton et al. (1976) and Avni (1976). This method is described in detail in *Numerical Recipes* (Press et al. 1992).

This chi-squared minimization technique encounters a variety of problems and ambiguities when treating real X-ray spectroscopy problems:

1. Binning requires an arbitrary choice of origin and bin width. Statistical techniques to choose optimal and adaptive binning (narrower bins where more data are present) are available (Silverman 1986), but are rarely used in astronomy.
2. Bins with few or zero counts may be arbitrarily deleted or used with *ad hoc* modified uncertainties.
3. The spectrum is usually background-subtracted. If either the source or background spectrum has low signal, then the error structure of the subtracted spectrum is neither Gaussian nor Poisson.
4. The degrees of freedom are typically 'bundled' to reduce the number of free parameters to fit. For instance, the abundances of different elements are collected into a single parameter scaled to solar abundances. The degrees of freedom are thus not uniquely defined for a given problem.
5. Similarly, the number of 'interesting' parameters is very subjective, shifting with the temporary goals of the researcher.
6. It is unclear whether the  $\chi^2_{min} + const$  confidence intervals are meaningful when  $\chi^2_{\nu,min}$  is below unity, as a wider range of models may be satisfactory. If  $\chi^2_{\nu,min} > 1$ , then no model is satisfactory. Thus the method may give meaningful confidence intervals only in limited cases.

In 1900, Pearson designed the statistic for a specific problem, the multinomial experiment (Babu & Feigelson 1996, pp. 64f). Its properties under the complicated non-linear parameter estimation conditions encountered in X-ray spectroscopy are not fully investigated. Several astronomers have suggested modifications to chi-squared minimization, typically by reweighting the denominator in equation (1) (e.g. Nolan et al. 1993; Wheaton et al. 1995; Kearns et al. 1995). While these modified- $\chi^2$  procedures do appear to perform better under some conditions, their mathematical properties (e.g. is the statistic  $\chi^2$  distributed?) have not been determined.

## 2.2 Possible alternative procedures

Perhaps the simplest approach is to convert the data and the models into normalized cumulative distributions and apply well-established nonparametric empirical distribution function (EDF) goodness-of-fit tests to establish whether the data are consistent with a given model. The EDF of an X-ray spectrum is a monotonically-increasing step function from 0 to 1 which rises by  $1/N$  at each energy a photon is detected, where  $N$  are the total photons observed from

the source (including background). It will look like a wiggly staircase, and is the only full nonparametric representation of a univariate dataset. The model contains both the complicated astrophysics (continua, emission and absorption lines) and a model of the detector background. Several statistics can be used to establish the probability that the data and a model are consistent (Daniel 1990; Babu & Feigelson, 1996, pp. 65ff):

**Kolmogorov-Smirnov** test is based on the supremum distance between the model and EDF. It will be most sensitive to large-scale differences in the spectra, such as the spectral index in a nonthermal spectrum.

**von Mises** test measures the sum of the square deviations between the model and EDF. It may be most sensitive to many small-scale deviations, such as the abundance of elements regulating the strength of emission lines.

**Anderson-Darling** is a weighted version of the von Mises test which increases sensitivity to deviations at the tails of the distribution. It may be most sensitive to the low energy absorption cutoff.

The astronomer provides a grid of possible spectral models, applies these tests to each data-model pair, obtaining for each a probability that the data and model are mutually consistent. The result can be viewed as a series of confidence level contours (90%, 99%, etc.) in parameter space showing acceptable models (e.g. Fasano et al. 1993). The method does not select a single model as the 'best-fit'. This nonparametric approach has many advantages over  $\chi^2$ -minimization: they require no binning, no minimum count rate, no counting of degrees of freedom or interesting parameters, and no assumption of error distributions. But they are often not the most powerful tests for discriminating differences between data and models.

A second promising approach is maximum likelihood estimation (MLE). Here a parametric model for the underlying process must be assumed. In X-ray astronomy, it is plausible to view the observed spectrum as a Poisson representation of the astrophysical model. The probability that a dataset is associated with a particular model can then be computed, and there exists a set of the model parameters  $\hat{\theta}_0$  that maximizes this probability. For binned data,

$$L(\hat{\theta}_0) = \max_{\bar{\theta}} \prod_{i=1}^k \left( \frac{M_i(\bar{\theta})^{N_i}}{N_i!} e^{-M_i(\bar{\theta})} \right). \quad (2)$$

Poisson MLE for nonspectroscopic problems in X-ray astronomy are discussed by Cash (1979), Sciortino & Micela (1992) and Wheaton et al. (1995). MLE has been a central pillar in applied statistics for several decades.

Finally, resampling is an important innovation that can assist with the estimation of confidence intervals (i.e. uncertainties) for spectral parameters for any of the above methods. Bootstrap resampling is a convenient computational tool that gives reliable probabilities for a majority of statistics (or confidence intervals in parametric models) for a very wide range of datasets. While many Monte Carlo simulation methods have been envisioned, mathematical study during the 1980s showed that the particular procedure used in the bootstrap provides very reliable confidence intervals. Its use is described by Efron & Tibshirani (1993) and Babu & Feigelson (1996, pp. 93ff).

A bootstrap resample of an X-ray spectrum with  $N$  photons has  $N$  randomly drawn photons, where the entire dataset is used for each drawing. The resample thus has multiples of some photons and is missing others. Each resample is then analyzed in the same manner as the original dataset (e.g. computing a statistic for a grid of possible models).  $O \sim N(\log N)^2$  resamples should be obtained, and a histogram of the best-fit parameter (e.g. plasma temperature or absorption column density) is constructed. Once a small bias is corrected, the confidence intervals for the parameter is immediately read off the histogram.

### 3 Improving Our Methodology

The need for statistical expertise and advances in modern astronomical research is clear. Observatories now generate terabytes of raw data, which are reduced to thousands of images or catalogs with  $10^5 - 10^9$  objects. Interpretation of data may entail parameter estimation for complex astrophysical models, treatment of biases from flux limits and measurement errors, or (deceptively complex) bivariate linear regressions for estimating the expansion rate of the Universe. About 500 papers/year have the word 'statistics' or 'statistical' in their title, and  $\sim 200$  of them are principally devoted to methodological issues falling at least partly under the purview of mathematical statistics. Though it may violate our self-conceptions, astronomers are frequently applied statisticians.

It is thus remarkable, and lamentable, that few of these studies reveal substantial awareness of the enormous progress in mathematical and applied statistics during the past few decades. Most astronomical studies overuse a few simple methods like  $\chi^2$  minimization and Kolmogorov-Smirnov tests, ignoring a vast range of methods reviewed, for example, in the 10-volume *Encyclopedia for Statistical Sciences* (Kotz et al. 1982-90) and the 12-volume *Handbook of Statistics* (Elsevier). The *Astrostatistics* monograph (Babu & Feigelson 1996) provides a brief overview and useful references for a number of statistical fields

such as exploratory data analysis, probability distributions, parameter estimation and confidence intervals, hypothesis testing and correlation, nonparametric tests, multivariate analysis, time series analysis, spatial statistics, wavelets, density estimation, sampling designs, and bootstrap resampling.

We can, however, expect the statistical community to reciprocate our attention and learn about astrostatistics meetings and initiated collaborations in astronomy. Agency and project managers should be aware that statisticians often work as consultants on projects in other fields. We thus must devote a small fraction of our funding towards them in order to obtain consistent, high-quality assistance and collaborations. Statisticians who take the time to understand our research problems can help astronomy at two levels: directing us to existing methodologies and, when these are inadequate, extending methodologies to meet our needs. Such an effort has begun at Penn State's *Statistical Consulting Center for Astronomy*. Queries arrive by email, straightforward answers are placed on the Web (<http://www.stat.psu.edu/scca>), and some problems requiring modest mathematical effort are developed in papers (Akritas & Siebert 1996; Akritas & Bershady 1996). Altogether we believe there is great potential for qualitative improvement in the analysis methodology, and thereby improvement in the scientific return from astronomical research.

### Acknowledgments

We thank Michael Akritas and Brice Lindsay (Statistics, Penn State) for informative discussions. This work was supported by grants NSF DMS-9626189 and NASA NAGW-2120.

### References

1. Akritas, M. G. and Bershady, M. A., *Ap. J.* **470**, 706 (1996).
2. Akritas, M. G. and Siebert, J., *M.N.R.A.S.* **278**, 919 (1996).
3. Avni, Y., *Ap. J.* **210**, 642 (1976).
4. Babu, G. J. and Feigelson, E. D., *Astrostatistics* (Chapman & Hall, London, 1996).
5. Babu, G. J. and Feigelson, E. D., *Statistical Challenges in Modern Astronomy II* (Springer, New York, 1997).
6. Cash, W., *Ap. J.* **288**, 939 (1979).
7. Daniel, W. W., *Applied Nonparametric Statistics* (PWS-KENT Publ., Boston, 1990).

8. Efron, B. and Tibshirani, R. J., *An Introduction to the Bootstrap* (Chapman & Hall, New York, 1993).
9. Fasano, G., Pisani, A., Vio, R. and Girardi, M., *Ap. J.* **416**, 546 (1993).
10. Feigelson, E. D. and Babu, G. J., *Statistical Challenges in Modern Astronomy* (Springer, New York, 1992).
11. Feigelson, E. D. and Nelson, P. I., *Ap. J.* **293**, 192 (1985).
12. Hald, A., *A History of Probability and Statistics and Their Applications before 1750* (Wiley, New York, 1990).
13. Jaschek, C. and Murtagh, F. (eds.), *Errors, Bias and Uncertainties in Astronomy* (Cambridge Univ., Cambridge, 1990).
14. Kearns, K., Primini, F. and Alexander, D., in *Astronomical Data Analysis Software and Systems V*, R. A. Shaw et al. eds. (Astro. Soc. Pacific, San Francisco, 1995).
15. Kotz, S., Johnson, N. L. and Read, C. B., *Encyclopedia of Statistical Sciences* (Wiley, New York, 1982-90).
16. Lampton, M., Margon, B. and Bowyer, S., *Ap. J.* **208**, 177 (1976).
17. LaValley, M. P., Isobe, T. and Feigelson, E. D., *BAAS (Software Reports)* **24**, 839 (1992).
18. Nolan, P. L. et al., *Ap. J.* **409**, 697 (1993).
19. Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P., *Numerical Recipes: The Art of Scientific Computing* (Cambridge Univ, Cambridge, 1992).
20. Rolfe, E. J. (ed.), *Statistical Methods in Astronomy*, ESA SP-201 (ESA Sci. Tech. Publ., Noordwijk, 1983).
21. Sciortino, S. and Micela, G., *Ap. J.* **388**, 595 (1992).
22. Sheynin, O. B., *Archive History of Exact Sciences* **11**, 97 (1973).
23. Silverman, B. W., *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, London, 1986).
24. Stigler, S., *The History of Statistics: The Measurement of Uncertainty Before 1900* (Harvard Univ., Cambridge, 1986).
25. Trumpler, R. J. and Weaver, H. F., *Statistical Astronomy* (Univ. California, Berkeley, 1953).
26. Wheaton, W. A. et al., *Ap. J.* **438**, 322 (1995).