

The Role of Statistics and Statisticians in the Future of Astrostatistics

Joseph M. Hilbe

T. Denney Sanford School of SFD, Arizona State University,
P.O. BOX 873701, Tempe, AZ 85287-3701
email: hilbe@asu.edu; j.m.hilbe@gmail.com

Solar System Ambassador Program, Jet Propulsion Laboratory,
4800 Oak Grove Dr., Pasadena, CA 91109

Abstract. Given the generic definition of statistics, it is clear that astronomers have engaged in statistical analysis of some variety since astronomy first emerged as a science. However, from the early nineteenth century until the beginning of the twenty-first the two disciplines have been somewhat estranged — there was no formal relationship between the two. This has now changed, as is evidenced by the recent creation of the International Astrostatistics Association (IAA), the ISI astrostatistics committee, astrostatistics working groups authorized by the IAU and AAS, and this Symposium. The challenge for us to come is in establishing how statisticians and astronomers relate in developing the discipline of astrostatistics. I shall propose a direction for how the discipline can progress in both the short term and well as for future generations of astrostatisticians.

Keywords. methods: statistical

1. Introduction

When I was first invited to present this Closing Address from my perspective as a statistician, I thought of a host of items that could be mentioned and discussed. However, when it came time to actually preparing my remarks for this occasion, I decided that it was best to focus on a single theme. This theme relates to the role that statistics and statisticians have in the future of astrostatistics. However, dual sub-themes are entailed in this general theme. I shall provide an outline of the recent events leading to the emergence of astrostatistics as a discipline in its own right, and second, I should like to propose a vision of the future of astrostatistics. Once I have proposed how I envision this future, I hope that the role statistician's play in it will be apparent.

2. Astrostatistics as a discipline

Astrostatistics may in general be defined as the discipline related to the statistical analysis of astronomical data. Some may incorporate astroinformatics as a component of astrostatistics, and a few years ago it was even argued that astrostatistics is a component of astroinformatics.

Astrostatistics focuses on statistically evaluating astronomical data; astroinformatics focuses on the problems and techniques of collecting and managing astronomical data. These are definitely related disciplines, and astrostatistics largely depends on, and will continue to rely on, the advances made in astroinformatics. However, the fields have essentially different goals, and I will separate the two for my purpose here.

Is astrostatistics a subdivision of the astronomical sciences? I tend to think that most astrophysicists believe this to be the case. Many statisticians, however, appear to think of astrostatistics as a branch of statistics, as they analogically do for biostatistics, geostatistics, environmetrics, chemometrics, psychometrics, econometrics, and epidemiology, to name a few such joint statistical disciplines.

Those in the biological and medical sciences with little exception think of biostatistics and biometrics as separate disciplines to biology. Geologists, though, I believe consider geostatistics as both a subdivision of geology and as a discipline in its own right. I envisage astrostatistics to be more in line with geostatistics in this regard, but the analogy to biostatistics may characterize the astrostatistics of the future.

Until recently astrostatistics has differed from these other joint statistical disciplines in that each of them have had an international association devoted to their discipline for a number of years. There are also either MS or PhD degree programs in these joint disciplines, and grad students can earn a degree with a clear concentration in the joint discipline.

A few examples of organizations that have been formed to support joint statistical disciplines include:

- the *International Environmetrics Society*,
- the *International Association for Mathematical Geology*,
- the *International Biometric Society*,
- the *International Society for Computational Biology*,
- the *International Epidemiology Association*.

A few examples of such organizations based within the social sciences include

- the *Econometric Society for the statistical analysis of economic data*,
- the *International Association for Business and Industrial Statistics*,
- the *International Association for Official Statistics*, and
- the *International Association of Survey Statisticians*, to name a few.

There are similar international associations for most every recognized joint statistical discipline I can think of. At the very minimum about every international scientific association I researched had a section or committee related to the statistical analysis of that discipline's data. Until four years ago astrostatistics had no international association, and only in August, 2012 — twenty-one months ago — did the IAA come into existence. Moreover, there are currently still no MS or PhD degree programs in astrostatistics of which I am aware, although a few universities are moving in that direction.

Astrostatistics, as a discipline, is still in its infancy. This is not to say that some type of statistical analysis, albeit primarily descriptive analyses, has not been applied to astronomical data for over two thousand years. It has. But except for a relatively few occasions, astronomy progressed perfectly well without the need for inferential statistical techniques beyond linear regression — until recently.

3. The recent development of astrostatistics

It was not until computers were powerful enough in terms of RAM, processor speed, and storage space, that more complex non-linear and non-parametric statistical procedures were developed. The advent of personal computers made interactive statistical analysis a possibility. With them came the rapid development of more complex statistical procedures, and as computing power increased, statistical capabilities increased. As more complex statistical procedures were developed, the more they could be of use to astronomers who desired to statistically analyze their data.

Although there were a few conferences between statisticians and astronomers prior to 1991, it was not until then that astronomers became more interested in statistics. The first of the quint-annual *Statistical Challenges in Modern Astronomy* conferences was developed by Pennsylvania State Professors Eric Feigelson in astronomy, and Jogesh Babu, in statistics. Sponsored by the Center for Astrostatistics at Penn State, these conferences have been the mainstay of the discipline until just recently.

Table 1. Recent Milestones in Astrostatistics (selected)

| | |
|----------|---|
| Jul 2008 | ISI Astrostatistics Interest Group formed (Hilbe) |
| Mar 2009 | Astro2010 State of the Profession Paper: Astroinformatics: A 21 st Century Approach to Astronomy (Kirk Borne, lead author) |
| Aug 2009 | Meeting of ISI Astrostatistics Interest Group, ISI World Statistics Congress, Durban, South Africa (Hilbe) |
| Dec 2009 | ISI Astrostatistics Committee approved (Hilbe) |
| Jan 2010 | ISI International Astrostatistics Network formed (Hilbe) |
| Aug 2011 | 2 Invited Papers and 2 Special Topics in astrostatistics sessions: (Sen; Hilbe) World Statistics Congress, Dublin, Ire – 1st invited sessions at WSC |
| Sep 2011 | Statistical Challenges in Modern Astronomy V (Penn. State- Feigelson, Babu); Springer Series in Astrostatistics formed at SCMAV conference (Hilbe) |
| Feb 2012 | ASAIP started (Feigelson & Hilbe editors) |
| Jun 2012 | AAS Astrostatistics & Astroinformatics Working Group approved (Ivezic) |
| Aug 2012 | IAU Astrostatistics & Astroinformatics Working Group approved (Feigelson) |
| Aug 2012 | IAA officially approved 30 Aug 2012 (Hilbe) |
| Jan 2013 | 1st meeting of the AAS A&A WG (at Winter meeting of the AAS), Long Beach, CA. (Ivezic) |
| Aug 2013 | Invited and Special topics sessions in astrostatistics at ISI World Statistics Congress, Hong Kong, China (Sen; Hilbe; van Dyk) |
| Mar 2014 | American Statistical Assoc. astrostatistics interest group approved (Cisewski) |
| Apr 2014 | IAA Working Group Cosmostatistics formed (de Souza) |
| May 2014 | 1st IAU Astrostatistics Symposium - Statistical Challenges in 21st Century Astronomy, IAUS306, Lisbon, Portugal (Heavens; Starck) |

Several collaborations of astronomers and statisticians were also developed during the 1990s and into the 21st century. But there were no committees or even working or interest groups in astrostatistics under the auspices of any astronomical or statistical organization or society until 2008.

Although I have a varied background, to be sure, including being an Emeritus Professor at the University of Hawaii and now adjunct professor of statistics at Arizona State University, I came to astrostatistics as a biostatistician having rather substantial experience in handling large national-level health-outcomes data. I also had spent considerable time programming statistical software procedures, a number of which became components of commercial statistical software.

As a youth, though, I had planned on becoming an astronomer, and only changed majors half-way through my undergraduate career in the mid-1960s. However, I still retained my interest in astronomy over the years, and as an avocation, collected meteorites. I have well over 100 meteorites of all varieties including Martian and Lunar stones and every classification of carbonaceous chondrite. In addition, I have been a member of the international Meteoritical Society for a number of years, which is mostly comprised of planetary scientists and geologists. In 2007 I became associated with the Jet Propulsion Laboratory, representing them in giving presentations on meteorites to schools, community groups, astronomy clubs, and the like. It was that association which resulted in my being here today.

In 2008 I was on one of my bi-weekly conference calls with the Principal Investigator of a JPL research study. I don't recall the nature of the project or mission now, or who the PI conducting the call happened to be, but I clearly recollect that he complained that he was not sure of how best to statistically evaluate study results. I had heard this same complaint before on several previous conference calls with other principal investigators, but this time the words stuck.

Not being aware at the time of the already ongoing collaborations in astrostatistics, I got off the call thinking that there should be a way to get astronomers and statisticians who have an interest in astronomical research together, from the very outset of a project,

in order to help affect a proper statistical analysis of the study data. I considered that astrophysicists have training and expertise on the physics side of such a project, but were likely not trained in statistical analysis — in particular only a relatively few are likely current with recent advances in statistics, but most would value having competent statistical advice. On the other hand, I thought that there would be a number of statisticians who would love to collaborate with astrophysicists on various astronomical research projects. It could be a win-win for both astronomers and statisticians, with better science as the result. I have amended my view a bit since 2008, but in large part I still maintain this position.

The call spurred me to form an interest group in astrostatistics under the auspices of the International Statistics Institute, or ISI, which is to the discipline of statistics as the IAU is to astronomy. I already chaired an ISI standing committee, and was acquainted with ISI administration and the procedures of becoming an interest group. Moreover, I soon learned that several very competent groups of statisticians and astronomers had been gathering for collaboration and conferences for a decade and a half, but I believed that more collaboration between astronomers and statisticians could exist if astrostatistics as a discipline was perhaps formally recognized by the ISI and IAU. Remember, I first conceived of this idea a result of hearing complaints about the need for statistical support from principal investigators who were associated with major research organizations.

The ISI Interest Group was formed for the purpose of gathering interested statisticians together at the 2009 ISI World Statistics Congress in Durban, South Africa to discuss astrostatistics and the role statisticians can play in the analysis of astronomical data. I convened the meeting, which was attended by some 50 astronomers and statisticians from around the world. The ISI president-elect also attended. At the end we voted to seek full committee status under the ISI. A full standing committee on astrostatistics was approved at the 2009 winter conference of the ISI executive committee — the first such committee under any astronomical or statistical organization.

I contacted those who I believed to be recognized at the time as the leaders in the discipline to join the committee. The response was thoroughly positive. The ISI, though, limited committee sizes to 15, and we had many more researchers interested in being involved. I therefore initiated the ISI Astrostatistics Network in January 2010, a few weeks after the committee had been approved.

In February of 2012 Eric Feigelson, also a member of the Network executive committee, was able to secure support from his department at Pennsylvania State University for the creation and maintenance of an Astrostatistics & Astroinformatics Portal, or ASAIP. He kindly invited me to be co-editor, a position we jointly share, but he has been the mainstay of the effort.

The ASAIP serves as a means for anyone interested in astrostatistics or astroinformatics to obtain information about forthcoming astrostatistically-related conferences, to receive announcements from other members, to learn about journal articles and books related to astrostatistics, and to even serve as a resource for job openings and job searches. There is also the opportunity for Portal members to join blogs on various topics.

For example, the ASAIP has 8 general purpose Sections of Blogs hosted by Portal members as well as three sections of Blogs that are primarily hosted by outside sites - 8 in astronomy, 15 in statistics, and 17 in information sciences. In addition to Blogs, the Portal provides information on locating tutorials and classes on subjects related to statistical modeling, software languages, and the like. The Portal had some 150 members by April 2012 and now has over 700 members, just two years later.

In June of 2012 the *American Astronomical Society* Council approved a new Working Group in Astrostatistics and Astroinformatics - the first such Working Group authorized

by an astronomical society. Shortly afterwards, in August 2012, Commission 5 of the IAU approved the creation of a similar Working Group, with Feigelson as its inaugural chair. Later that month, on the 30th of August, the executive committee of the ISI

Astrostatistics Network voted to be renamed the *International Astrostatistics Association*, or IAA. It is now an independent professional association for astrostatisticians with ties to the ISI through the ISI astrostatistics committee, and to the IAU through its relationship to the IAU Working Group in Astrostatistics & Astroinformatics.

Currently the IAA has some 462 members from 47 nations, with membership sections of regular, PostDoc and grad student. Regular members are also partitioned as being astronomers, statisticians, information scientists or astroinformaticists, or other. Last month the IAA inaugurated a Working Group on Cosmostatistics aimed at promoting the development of new statistical tools for cosmological data exploration. Rafael de Souza of the Korea Astronomy and Space Science Institute initiated and heads the working group. As of May 15th, there were 31 members participating in the Working Group, with the four project groups displayed on the screen already having been formed. A fifth project group was formed this past week.

*General Relationships into Astronomical Datasets,
Multiple Survey Analysis,
CosmoR : R oriented packages for cosmostatistics,
Generalized Linear Models for Astro/Cosmo-Statistics.*

4. Astrostatistics and extended generalized linear models

With respect to the last mentioned Project Group—on Generalized Linear Models or GLMs—I believe that astronomers have underutilized this class of models and their extensions. Basic GLMs include normal or Gaussian regression, gamma and inverse Gaussian models, and the discrete response binomial, Poisson and negative binomial models.

Table 2. Range of Generalized Linear Models

| |
|---|
| Continuous Response Models |
| Normal or Gaussian, Log-normal |
| Gamma |
| Inverse Gaussian |
| Binomial Response Models |
| Logistic - probit - complementary loglog - loglog |
| Count Response Models |
| Poisson (POI) |
| Negative Binomial (NB) |

Extended Generalized Linear Models include the addition of scale parameters for continuous response GLM models, as well as a wide variety of discrete response regression models. Many astrostatistical data situations have discrete data to be modeled, but are instead modeled as if the variable of interest to be modeled is continuous. Or, if they model count data, many astronomers employ a Poisson model, somehow forgetting about the Poisson distributional assumption of the equality of the mean and variance—called equi-dispersion. This is a terribly restrictive assumption, and is rarely found in real data.

From looking at the literature in the field, it appears that most astronomers are not aware of the explosion of discrete response models that have occurred recently—many of which I believe can be applied to astronomical data from both frequency-based and Bayesian perspectives. Project Groups like this—comprised of both astronomers and

statisticians — have the prospect of bringing more modeling examples, and software, to the attention of astronomers seeking to properly evaluate their study data.

Table 3. Selected Extensions to the Basic GLM Algorithm

| | |
|---|--------------------------------------|
| Binomial | |
| Beta | Generalized maximum entropy logistic |
| Generalized binomial | |
| Nested logit | Fractional polynomial binomial |
| Proportional odds models | Partial proportional odds models |
| Multinomial models | Heterogeneous multinomial models |
| Beta-binomial | |
| Count | |
| Heterogeneous negative binomial | Poisson-lognormal |
| Poisson Inverse Gaussian (PIG) | 3-parameter generalized NB-F |
| Generalized Poisson (GP) | 3-parameter generalized NB-W |
| Generalized negative binomial P | Finite mixture models |
| Double Poisson | Quantile count models |
| Double hurdle | |
| 2-part Hurdle POI, NB, PIG, GP, NBP, NBF & Logit, Probit, POI, etc | |
| Zero-inflated POI, NB, PIG, GP, NBP, NBF & Logit, Probit, cloglog, loglog | |
| Truncated POI, NB, PIG, GP, NBP, NBF (right, left, interval, 3+levels) | |
| Censored POI, NB, PIG, GP, NBP, NBF (right, left, interval, 3+levels) | |
| Hierarchical and mixed standard and Bayesian models for the above | |

5. The future of astrostatistics

I hope that we can encourage more participation in the IAA *Working Group in Cosmostatistics*, with other astrostatisticians proposing or becoming members of similar project groups. New working groups can be developed as well. To my mind this represents a true advance in the discipline, and our discipline will thrive as this program expands.

Regarding my themes — the role of statistics and of statisticians in the future of astrostatistics: I have argued that both statisticians and astronomers play essential roles in the discipline. Statisticians will continue to develop ever more complex statistical methods, many of which will be applicable to astronomical science. In fact, recent advances in environmental and ecological statistics, which entail work in spatial analysis, geostatistics, and discrete response modeling, and advances made by econometricians in time series, can be of particular use to astrostatisticians. A number of other newly developed statistical methods can as well.

Statisticians who are working in these areas, as well as competent statisticians who have a particular interest in astronomy, can be of considerable help to astronomers when analyzing data.

Of course, there may be some astronomers who love to engage in statistics and perhaps thoroughly enjoy developing statistical models and software appropriate for astronomical analysis. Those attending this Symposium give evidence of this type of astrophysicist. You have the capability of developing statistical procedures and software that not only is applicable to astronomy, but which can also favorably impact research in other disciplines.

Table 4. Proposals for the Future

- Joint Astrostatistics MS & PhD Degree Programs
- Having more researchers propose and develop IAA Working and Project Groups
- Structure IAA to be like similar international subject area associations
 - § Add committees to IAA
 - Committee for PostDocs
 - Committee for Graduate Students
 - Membership committee
 - Awards-Recognition committee
 - For outstanding publication award
 - Young researcher awards
 - IAA Fellow
 - Signature award for contributions to the discipline
 - Meetings committee
 - Coordinate and advise members of the various meetings and conferences related to astrostatistics & astroinformatics being held throughout the year
 - Finance committee
 - I have sought to have external funding support the activities of the IAA. I hope we can maintain this philosophy. We need more than one person working on this task.
 - § Other functions ???

I also propose that we encourage and participate in the development of interdisciplinary degrees in astrostatistics, from the MS to the PhD level. These should be joint projects by University Astronomy/Astrophysics and statistics departments. Such programs will entail the need for participating statistics departments to hire professors who specialize in astrostatistics. This is exactly what occurs in other joint disciplines, and will in astrostatistics as well. This is good for the discipline, and will help increase job opportunities for astrostatisticians.

Finally, I believe it is important to develop and support the IAA, as well as the IAU and AAS Working Groups. But if our discipline is to advance and flourish, and be thought of as are other successful joint disciplines, maintaining and supporting a strong international astrostatistics association is vital. It has been for other joint disciplines, and it will be for us as well.

As a statistician who has had a lifelong interest in astronomy and cosmology, I have been pleased to participate in this IAU Astrostatistics symposium and to serve on its scientific organizing committee. The outstanding success of this Symposium should signal the IAU that astrostatistics is a valuable component of professional astronomy, as well as an important discipline in its own right.

It's now up to us, both individually and collectively, as astronomers, statisticians and information scientists, to decide the future direction of astrostatistics, and how we wish to shape its future.

Reference

Hilbe, J. M. 2014, *Modeling Count Data*, Cambridge: Cambridge University Press