UC Irvine

UC Irvine Previously Published Works

Title

Late-night thoughts of a classical astronomer

Permalink

<https://escholarship.org/uc/item/9t6644r2>

Author

Trimble, V

Publication Date

1997

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at

<https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Late-Night Thoughts of a Classical Astronomer

Virginia Trimble¹

ABSTRACT Conference participants from the astronomical side of the fence were astounded by the wide range of statistical concepts and techniques available, at least in principle, for use. Complimentarily, those from the statistical side expressed surprise at the enormous range of kinds of astronomical data (in spatial, temporal, wavelength, and other domains) crying out for more sophisticated analysis. Nevertheless, bringing the two together is not to be done over afternoon tea, and real collaborations, not just advisors, are needed.

21.1 Introduction: The astronomical data base

I arrived at Penn State with a prediction at hand: “Nobody is going to learn anything at this meeting.” That is, I expected that the astronomers would not actually be able to carry out any analyses that they hadn’t been able to do before, and that the statisticians would not be able to take home any specific databases and do anything with them. I believe this turned out to be at least roughly the case. What then was the purpose in coming? To find out what is out there, in terms of methods and problems and, perhaps more important, who is out there who knows something about the techniques or the data that one might be interested in. One of the speakers reminded us that statisticians should not be regarded as shoe salesmen, who might or might not have something that fits. But at least some of the leather merchants have had a chance to meet some of the shoemakers.

The rawest possible astronomical information consists of individual photons (or the radio-wavelength equivalent in Stokes parameters) labeled by energy, time of arrival, two-dimensional direction of arrival, and (occasionally) polarization. Such photons may come from the sources (stars, galaxies, ...) you are interested in, but also from unwanted sky, telescope, and detec-

¹Physics Department, University of California, Irvine CA 92697-4575 and Astronomy Department, University of Maryland, College Park MD 20742

tor backgrounds, each also with its own temporal, spatial, and wavelength patterns, often not well described as standard Gaussian noise.

Sources, telescopes, detectors, and backgrounds are very different beasts in the different wavelength regimes: radio, microwave, infrared, visible, ultraviolet, X-ray, and gamma-ray. For instance, both spatial and energy resolution are generally poor for X- and gamma-rays, but you can record photon arrival times to a microsecond or better. Radio interferometry buys you enormous spatial resolution, but at the cost of losing extended emission. Ground-based work in the infrared is always background-limited, because air, ground, telescope, and all the rest are 300 K emitters. Excellent wavelength resolution is possible for visible light, but only if you are prepared to integrate the photons from faint sources for many hours. Unless “source” photons greatly outnumber “background” photons, statistical care is already needed at the raw data level to decide whether there is a source real enough to worry about, whether it is compact or extended, and where it is in the sky.

The next level of processing takes the individual photons coming from some part of the sky and assembles them into spectra (number of photons or brightness *vs.* wavelength during a given time interval), light curves (brightness *vs.* time in a fixed wavelength band, that is, time series), or maps (brightness *vs.* position in the sky during a given time interval and in a fixed wavelength band, also called images or pictures). At this stage, one asks statistical questions about the reality of particular emission or absorption features in spectra, about whether a source is truly variable and whether periodicity can be found in the light curve, and about the reality and morphology of apparent features in the maps. An important question at this level is whether adding another parameter to your fit (emission line, pulsation mode, subcluster, ...) improves the fit enough to be believable. Astronomers have historically used the chi-squared test for this purpose, and indeed are typically not aware of any alternatives. One virtue of Bayesian methods is that they automatically vote against extra parameters for you unless the addition makes an enormous improvement.

Stage three compares processes spectra, light curves, maps, etc. with previously-existing templates and assigned sources to known classes like peculiar A stars (ones with surface temperatures near 10,000 K and anomalously strong absorption lines of europium and related elements) or BL Lac objects (active galactic nuclei characterized by rapid variability and weak emission lines). Often data from more than one wavelength band must be combined at this stage to recognize classes like low-mass X-ray binaries (pairs of stars with one neutron star or black hole accreting hot gas from a solar-type companion), gamma-ray pulsars (which show the same rotation period at radio and gamma-ray wavelengths), or Fanaroff-Riley type 2 radio galaxies (ones whose radio emission is very bright and has a particular double-lobed morphology). The main statistical questions at this stage concern whether the chosen template is a “good enough” match to

the new object (given the noise level), would another template be better, or have you found a genuine new class of source? In this last case, the two most immediate goals are (a) publish in *Nature* and (b) find two more, at which point, with three examples, you have discovered a well-known class of astrophysical object.

At stage four, we are ready to begin carrying out serious astronomical research, looking for correlations of properties among objects within and between template-defined classes in spatial, temporal, or other domains. Considerable theoretical firepower must be brought to bear at this stage. For instance, (a) given the distribution of brightnesses and colors of stars in a nearby galaxy, a luminosity-mass relation, and calculated lifetimes of stars as a function of mass, figure out the history of star formation $N(M, t)$ in the galaxy; (b) given the range of galaxy types and the current chemical composition of the X-ray emitting gas in a cluster of galaxies, figure out the history of nucleosynthesis in the cluster; or (c) given the sky positions and redshifts of a large number of galaxies, decide whether the pattern could have arisen from a particular spectrum of perturbations in a universe dominated by cold dark matter. Noise from all the previous levels is, of course, still with us, and the statistical nature of many conclusions shows in their being expressed in the form “we can exclude constant star formation rate (or dominance of Type Ia supernovae, or mixed dark matter) at the 95% confidence level”.

Invited review talks addressed problems at all of these levels, from deciding whether even one photon in a box belongs to your source (Jefferys, Ch. 3) to the topology of very large scale cosmic structure (Martinez, Ch. 9).

21.2 Astronomers on their own

The astronomer on the street is likely to be of the opinion (a) that statistics began with Gauss and (b) that the one thing we all understand is the least squares method. Prof. Rao’s (Ch. 1) historical introduction quickly disabused participants of both of these illusions. Soon thereafter, we became aware that the first two associations that the word “statistics” triggers in many of our minds (statistical parallax and Bose-Einstein, Fermi-Dirac, etc. statistics) were not terribly relevant to the purposes of the meeting, though the Luri et al. poster (Ch. 39) dealt with statistical parallaxes.

As always nomenclature (or jargon if you want to be insulting), including mathematical notation, was a significant part of the communications barrier between the two communities participating. While reading the poster presentations, I started making a list of the terms that I was not prepared to define accurately. When the total reached 99 (Table 21.1), I stopped, keeping in mind the Islamic tale that The Almighty has 100 names, of which

99 are known to mankind. The camel is so smug because only he knows the 100th name. At this point I encountered the word heteroscedastic, and can say only that the camel deserves it. Undoubtedly a comparable set of mysterious phrases from the astronomical dictionary could have been compiled with equal ease. Always mysterious are the eponyms (ideas, classes, methods, etc. named for a discoverer). Some terms seem to have an obvious meaning that is clearly wrong, like adaptive regularization (marrying your mistress before she shoots you?), and oriented pyramids (toward the east?). Others sound remarkably oxymoronic from outside (decision tree, annealing simplex method).

Astronomers, we have told ourselves repeatedly and have been told by other sorts of scientists and mathematicians, are particularly poorly equipped with statistical tools and careless in the use of the ones we have. I think there is some truth in this – scanning of several weeks issues of the *New England Journal of Medicine* yielded about twice as many separate methods and concepts as a couple of issues of the thrice-monthly *Astrophysical Journal* (which is **much** thicker). And, apart from Student's t-test and a couple of rank, regression, and correlation analyses, there were not the same methods either.

This is not to say that we are totally hopeless when left to our own devices. Eponymous examples include Malmquist (1920, 1924, who had a bias), Scott (1956, who had an effect), and Lutz & Kelker (1973, who had a correction). Malmquist noted that flux-limited samples (i.e. ones censored/truncated by the number of photons you can gather from a source in the allowed time) will always be deficient in intrinsically faint objects far away. Scott noted that, only when you have really big samples which requires looking at distant objects, will you find the rarest (including brightest) examples of a class. Lutz and Kelker pointed out that there are many distant stars with small parallaxes and few nearby ones with large parallaxes. Thus even symmetric errors will make it look like there are more nearby stars than is really the case. Errors are asymmetric in fact (since no parallax can be negative), making it worse. All three of these items deceive you into thinking that a distant set of stars, galaxies, etc. is closer than it is, and the first remains a major unresolved issue in establishing the extragalactic distance scale. If there are three cosmologists in the room, at least one will believe that at least one other does not understand Malmquist bias.

All three can be thought of as examples of truncated or censored data, many different approaches to which were discussed by Yang (Ch. 5), Segal (Ch. 4), Jefferys (Ch. 3), Duari, and Caditz (Ch. 38), but none is usually thought of as a statistical issue by practicing astronomers. We see them as akin to physical effects, like interstellar absorption, which you ignore at your peril.

Three more obviously statistical items were mentioned during the meeting that seem also to have been first thought of within the astronomical

TABLE 21.1. A subset of unfamiliar statistical terminology

hyperparameter	entropy prior
segmentation	Kahonen net
neural network	non-parametric
correlograms	differential distribution
gapped time sequence	siftware
multifractal	flatness spectra
adaptive regularization	Holden experiment
Slepian sequence	Gibbs sampler
à trous algorithm	embedding dimension
scalogram	scalegram
noise grams	errors of realization
maximum entropy methods	periodogram
autoregressive	multiscale entropy
pyramid, Gaussian	intermittency
pyramid, Burt	random coefficient
pyramid, Laplacian	Lebesgue density
pyramid, gradient	Fourth Minkowski functional
pyramidal vision	vision models
pivotal	i.i.d.
Huber estimator	intrinsic Bayes factor
minimum distance estimator	downhill simplex method
genetic program	annealing simplex method
multiscale edge detector	Lipschitz exponent
Lyapounov exponent	Sterrs
Gabor transform	gaussian window function
Boniferroni inequality	nested Lampton method
Ruger inequality	linear least squares estimator
gamma kernel	normalized fluctuation spectrum
Mallat's algorithm	Sobolev constraint
Wiener interpolator	Sterne-Campbell contingency test
binomial assumption	second-order intensity function
pursuit methods	Markov chain
supervised classification	Poisson regression
conjugate gradient algorithm	outliers
wavelet	axis-parallel decision tree
mother wavelet	K-function
Morelet wavelet	C-function
Mexican hat wavelet	Occam function
Haar wavelet	posterior space
Daubechies wavelet	Greenwood's formula
self-organized mapping	probability map
cumulative hazard rate	Heaviside's step function
B-spline function	Poisson bias
censored	proper weighting function
truncation	regressogram method
Lynden-Bell estimator	fuzzy correspondence analysis
Epanechnikov kernel	basis pursuit
linear state scale model	matching pursuit
Metropolis-Hastings algorithm	Wilcoxon statistic
Fisher matrix	

community. Historically first is $\log N - \log S$ (Ryle 1955), a plot of number of sources with brightness equal to or greater than flux S . Historically, it was used to show that radio galaxies and quasars have changed their average properties over billions of years. It is an integral method, with the advantage of not introducing binning noise but the disadvantage of propagating errors forward through the whole distribution.

$P(D)$ (Scheuer 1957) means “probability of deviation” and comes from the language of radio astronomy. It has been independently (re)discovered in the optical community as the method of surface brightness fluctuations for measuring distances to galaxies. The idea is that, when you try to count faint things (stars, radio sources, etc.) there get to be so many that you cannot resolve them as individual objects. But you can still say something about the number as a function of apparent brightness by looking hard at the fluctuations across the sky of their summed brightness, assuming that positions are random and the number of objects per resolution element is therefore describable by Poisson statistics.

Finally, V/V_m (Schmidt 1965) is a way of learning about how some sort of astronomical object is distributed in space even though your data sample is truncated in two or more dimensions (radio flux and optical brightness for the original quasar case that Schmidt considered). A modification can be used to construct luminosity functions from flux-limited data. Both V/V_m and $\log N - \log S$ have been used more recently to rule out certain possible models of gamma ray bursters.

21.3 Some common problems

This section includes most of the specific astrophysical issues discussed at the symposium, classified (very crudely) in parallel with the phases of data analysis mentioned in §1. Many sections are introduced by quotations from the conference participants and other players in the statistical arena. The content of the conference occupied (at least) a two-dimensional surface, where the y axis = statistical techniques and the x axis = astronomical problems. This summary is necessarily one-dimensional, and the “across-the-rows-and-down-the-columns” format of §3–4 results in many presentations being mentioned twice.

21.3.1 Source detection and image restoration

“Most of the bins have zero [source] photons in them, which is difficult even for statistics.” W. Jefferys

Indeed, a very common problem is one of deciding about the reality of a source (spectral line, etc.) seen against a non-randomly noisy background. The Theiler & Bloch (Ch. 30) and Damiani et al. (Ch. 33) posters addressed

precisely this issue for X-ray sources, the former using a new interpolation technique and the latter wavelet transforms.

A closely related task is that of cleaning and restoring images that have been messed up by processes of known (non-Gaussian) properties. The Hubble Space Telescope, with its improperly figured main mirror, has driven many recent efforts in this area. The Núñez & Llacer (Ch. 28) poster considered Bayesian methods, the Starck & Pantin poster (Ch. 29) multiscale maximum entropy methods, and the Anderson & Langer poster an existing package of pyramid and wavelet methods for image reconstruction. Wavelets are particularly appropriate for noise suppression (one form of which is called flat-fielding), and both Murtagh's talk (Ch. 7) and Kashyap's poster (Ch. 34) described wavelet approaches to simultaneous noise removal and source identification.

21.3.2 *Image (etc.) classification*

"If you have a spectrum, you immediately know if it's a star or a galaxy."
R. White (This is funny only if you know about cases like BL Lac, whose name says that it was first classified as a variable star and only later as the prototype of a subtype of active galactic nuclei, because its spectrum and light curve, considered in isolation, were every bit as ambiguous as its image.)

OK, you have decided you have a real source, and you are reasonably certain (perhaps even rightly so) that it belongs to one of a small number of discrete classes. How can you carry out the classification efficiently and automatically? Three problems of this sort were addressed. In each case, one starts with some sort of training set of images or spectra known to be correctly classified and an algorithm for deciding which class a new example belongs to (neural network, nearest neighbor, decision tree, matrix multiplication, rank order, or many others). The choice of critical parameters for distinguishing classes can be made by the programmer or by some of the types of programs.

White (Ch. 8) considered the most basic distinction — is an image that of a star or a galaxy (or plate flaw or cosmic ray hit)? — and a number of ways of selecting training sets and classification parameters (e.g. the rank-order of ratio of central to total brightness of an image in a field as a star/galaxy criterion). The Nail poster dealt with the next stage of classification, separating galaxies into normal and peculiar. Closely related is a recent exercise coordinated by Opher Lahav at Cambridge [*Science* 267, p.859, 1995], in which six senior extragalactic astronomers classified a number of galaxy images on a fine grid of subtypes, and then an artificial neural network did the same. It was roughly a tie. The Bailer-Jones poster (Ch. 42) presented a fairly preliminary attempt to automated classification of stellar spectra with a neural network. The "official" MK types are defined primarily by a small set of examples and secondarily by a small number of

specific line strength ratios. It was not clear whether the network had access to these official rules or was expected to find its own from the training set. The difficulty it found in separating type IV (subgiant) stars from main sequence and giant ones suggests the latter.

21.3.3 *Pattern recognition and description*

“What would you think if you looked at the 11 brightest constellations and saw 7 Big Dippers?” H. C. Arp, 1965 (in connection with arguments for non-cosmological redshifts of quasars). *“Well, you do see three.”* J. E. Gunn, 1965 (or even four – the Big and Little Dippers, Pleiades, and great square of Pegasus).

The topics mentioned here are distinct from those of §3.2 in that one does not start with a small set of *a priori* templates, but rather is trying to decide whether there is any interesting structure (in an image, spectrogram, time series, or whatever) and, if so, how should it be described.

The Lawrence et al. (Ch. 35) and Turmon (Ch. 31) posters both considered the patterns of magnetic activity on the solar surface from this point of view, using different statistical methods (wavelets, multifractals, and Markov random fields), but coming to rather similar conclusions, that the patterns are too complex for either verbal description or modelling to be very successful. The length scales on the Sun range from a little less than 1000 to about 1,000,000 km. A very similar problem, but on a scale of 10^{19} – 10^{22} km, arises when you attempt to characterize the large scale distribution of galaxies in the universe in terms of clusters, voids, filaments, sheets, or whatever. Betancort-Rijo’s poster (Ch. 26) considered this (though the inclusion of the phrase “random fields” in the title leaves me puzzled), as did the talk by Martínez (Ch. 9). One well-defined question is the basic topology of the large scale structure. Are clusters studded through spade like meatballs, voids scattered through denser regions like holes in swiss cheese, or something else? The answer, according to Martínez, seems to be a sponge-like topology, so that both over-dense and under-dense regions are connected up.

Pattern recognition is not, of course, a problem unique to astronomy, but has numerous applications in military reconnaissance, machine reading of handwriting, and so forth. Many of the methods developed in these areas, however, make heavy use of edges in the field, not very appropriate for astronomy, most of whose structure dribble off from dense cores to tenuous envelopes.

Whatever data set and methodology is under consideration, one has, as van der Klis (Ch. 18) reminded us, to worry about two different kinds of errors, equally serious. The first is seeing patterns that aren’t really there. Martian canals and other examples have made astronomers sensitive to these. Not seeing something that really is there is less embarrassing, but just as likely to impede physical understanding. These are called type 1

and type 2 errors, and, like Faranoff-Riley type 1 and type 2 radio galaxies, I forget which is which. But, in case you should ever need to know, Type II supernovae are found among population I stars, and type 1 Seyfert galaxies have broader emission lines than type 2's.

21.3.4 *Fitting known functions*

“The need for statistics arises because nothing in life is certain except death and taxes, unfortunately not in that order.” Anon.

Function fitting is the quintessential problem in statistical astronomy, since we were all taught that Gauss invented least squares in order to reduce a large number of observations of the first asteroid, Ceres, to a single best orbit. Newton's laws guarantee that (apart from perturbations due to Jupiter and such) the orbit can be described with a handful of numbers for semi-major axis, eccentricity, inclination of the orbit to the plane of the ecliptic, longitude of perihelion, and time of perihelion passage. To generalize from the solar system to any other pair of point masses you add the two masses as numbers 6 and 7. Dikova's poster (Ch. 48) concerned an outgrowth of Gauss's problem, identifying groups of asteroids with similar orbit parameters, while Ruymaekers & Cuypers (Ch. 40) considered the case of the reliability of the orbits of binary stars, but using a bootstrap method, not least squares. Acuna & Horowitz showed that you can fool some of the photons some of the time. If the image of a single point source seen with a given telescope/detector combination is very accurately known, then you can reliably recognize and assign brightnesses to two point source images even when their angular separation is somewhat less than the traditional Rayleigh criterion.

A great many other astronomical tasks are also of this general form, because one knows in advance, for instance, that the widths of magnetically broadened hydrogen lines will be linear in field strength, while their centroids shift in proportion to the square of the field; rotational broadening is linear in rotation speed, while turbulent broadening scales with the square root of the masses of the atoms responsible for the line; and so forth.

21.3.5 *Fitting unknown function, additional parameters, and goodness of fit*

“With five parameters, you can fit an elephant.” George Gamow (attributed)

Siemiginowska et al. (Ch. 14) asked a number of critical questions in this area, drawing from examples in X-ray astronomy, though they apply at all wavelengths. How do you find your way through many-dimensional parameter space to a “best fit”? Support the template you are trying to fit itself has uncertainties (e.g. in atomic data for spectral lines); how can

this be included in error estimates in Bayesian and frequentist methods? What should replace chi-squared as a test of goodness of fit when most of the information is in a few data points? (The Wilcoxon test was mentioned by several speakers as being suitable for picking out regions of largest deviation.) And so forth. Wheaton's poster (Ch. 41) suggested an alternative weighting scheme to be used in place of the standard chi-squared one for the specific case of bins with very few photons in them.

A common astronomical approach to the problem of finding relationships when you don't know what to expect in advance goes back to the early days of radio astronomy and says that, when you don't understand a phenomenon, the first thing to do is to plot it on log-log paper. Indeed, plots of log this *vs.* log that quite often yield clusters, correlations, and principal planes (for instance the Tully-Fisher, $D_n - \sigma$, Faber-Jackson, and Fundamental Plane methods of measuring distances to galaxies). The disadvantage is that one is often left with relationships that are not in any way physically understood. Mukherjee & Kembhavi's poster on the decay of pulsar magnetic fields considers a problem of this type. I believe that Qian's poster presents a relevant methodology, but found it difficult to interpret.

Other posters that addressed looking for correlations, structures, etc. that are not understood in advance include those of Starck & Pankin, Luri, Arenou et al., Chernenko, Csabai & Szalay, Betancort-Rijo (*cf.* Ch. 29, 39, 46, 47, 43, 26), and Schaefer et al. (not actually posted, though represented in the abstract booklet, and perhaps containing the last Fourier transform in the world). Most of these are also mentioned later under types of tests.

By the way, you really can fit an elephant with five parameters. Or, at least, the contour of the upper outline from trunk to tail tip can be traced fairly well with portions of five sine waves of different wavelength and amplitude.

21.3.6 *Time series*

"Yes, I can see the canals." J. P. Ostriker contemplating one of the results of J. D. Scargle's 1969 thesis analysis of periodic motions of wisps in the Crab Nebula.

Time series (or light curves of various kinds) was the astrophysical problem addressed in the largest number of talks and posters. It includes finding periods and power spectra, especially for data with large gaps, irregular intervals, or intervals near an actual period (leading to aliases). Because the astronomical community is small (and there are lots of stars), the loss of one productive observer can create such a data set from what would otherwise have been a much more satisfactory one. We can call this the 'Grant Foster effect', since he mentioned it during a discussion section. I had previously thought of it in connection with the Crab Nebula as "and then Baade died and stopped taking plates so often".

The intuitive method of period fitting has had some spectacular failures. In the 1930's for instance, Harlow Shapley found a number of variable stars in the Small Magellanic Cloud (SMC) that he thought should be RR Lyrae variables. And indeed with a few dozen irregularly spaced observations per star, he was able to fit every one of them with a reasonable RR Lyrae period (0.2 to 0.8 days or thereabouts). This implied a small distance to the SMC consistent with his previous ideas based on Cepheid (brighter) variables, and so helped to preserve a wrong extra-galactic distance scale for about 20 more years, until David Thackeray finally found the real SMC RR Lyraes in 1952, a factor of four fainter than Shapley's stars. Subsequent studies showed that each of Shapley's variables had a real period longer than 1 day, implying that the stars are larger and brighter than RR Lyraes and the SMC much further away than he thought.

A number of posters focussed on time series analysis for particular objects or classes, including Vityazav (Ch. 25) (correlograms as a problem in aperture synthesis), Lawrence et al. (temporal patterns of solar activity with wavelets and multifractals, Ch. 35), Lehto (light curve of the gravitationally lensed quasar OJ 287 with Haar wavelets, Ch. 36), Koenig (X-ray variability of the active galaxy NGC 5506 with linear state space models, Ch. 24), Young et al. (gamma ray bursts with wavelets, Ch. 37), Schaefer & Bond (quasi-periodic oscillations in AM Her binaries with Gabor transforms – like any eponym, this one invites the question, *Eva*, *Magda*, or *Zaza*?), Nair & Principe (long and short term predictions with neural networks for 3C 345 and PG 1159 stars), Hertz (reality of orbit period changes in X-ray binaries using several tests, Ch. 23), and Ringwald et al. (possible long-term variability of unknown origin in the cataclysmic variable GK Per; these data are nearly continuous and uncensored, but the authors still feel some residual uncertainty about the reality of the hecto-day oscillations, which were not at all expected).

Five of the oral presentations also concerned time series analysis for a variety of objects using a variety of methods. Van Leeuwen (Ch. 15) was concerned with the problem of enormous gaps in a time series in connection with trying to characterize variable stars from the HIPPARCHOS data base (not, of course, the primary purpose of this astrometric mission). This has indeed proven possible in a somewhat limited way (though the inventories are small compared, for instance, to the variable star byproducts of MACHO and OGLE searches for gravitational lensing). But his primary concern was with later astrometry missions and how they might be designed to provide more useful light curves without compromising the primary goals.

Guégan (Ch. 17) addressed embedding dimensions, Lyapunov exponents, and other ways of recognizing chaos (meant as a technical term, not a description of the normal state of astronomical research). A strictly positive Lyapunov exponent, for instance, means that two trajectories starting with arbitrarily similar initial conditions can eventually wander arbitrarily far

from each other. She noted that astronomy has one advantage over other disciplines (e.g. economics) in which chaotic behavior has been sought in having at least one authentic example: the tumbling rotation of Saturn's moon Hyperion (the analysis of which makes use of data collected as far back as the mid 1920's).

As Thomson pointed out, there is more than one way to handle gapped time series data. One set of methods looks only at the actual data points (e.g. the Lomb-Scargle periodogram), while the other begins by interpolating to fill in the gaps. Thomson favors interpolation methods and showed examples of their application to the light curves of the two main components of the gravitationally lensed quasar OJ 287 and to variability in the solar wind as charted by the Ulysses "over the pole" mission. The former bears on the value of the Hubble constant, the latter on the existence and properties of certain solar oscillations. Swank and van der Klis (Ch. 18) both addressed X-ray light curves, especially the ones with extremely fine time resolution now coming from the Rossi X-ray Timing Explorer (XTE or RXTE) and eventually to come from the Advanced X-ray Astrophysics Facility (AXAF). A wide range of high frequency (up to kHz) phenomena are turning up, and one needs to be able to answer questions like how to identify the shortest time scale present, how to set upper limits in power spectra, and how to characterize red noise in the presence of low frequency leakage, before the "botany" stage of studying X-ray binaries and other sources can yield astrophysical understanding.

21.3.7 Recognizing rare events and new classes of sources

"If you believe something one sigma is enough; if you don't, then 15 sigma won't help." Richard P. Feynman c. 1975

A surprising number of astronomers have worked over the years on what many of their colleagues believed to be empty data sets. Examples include the first ten years of extra-solar-system gamma ray astronomy (until the first source was found), SETI (the search for extraterrestrial intelligence), brown dwarfs (until about a year ago), the extragalactic infrared background, and gravitational radiation (known to exist and do what Einstein's equations predict, but through indirect evidence on binary neutron stars, not through direct detection). In terms of the phases of astronomical data processing described in §1, these are problems in which the first task is to create a template with which candidate events or sources can be compared.

Axelrod (Ch. 12) discussed the searches for gravitational lensing of stars in the Large Magellanic Cloud and the galactic bulge by "MASSive Compact Halo Objects" or MACHOS (deliberately coined to provide a contrast with Weakly Interacting Massive Particles or WIMPs). Potential lenses include known stars, substellar objects, planets, or anything else in the disk or halo of our own galaxy with masses in the range 10^{-3} to 10 solar masses. The team began with a template for a MACHO event that required the

brightening and fading to be time symmetric, colorless, and other various definite things. Actual observations and further thought have required the template to evolve and include lensing of and by binary and moving stars (not time symmetric) and lensing of stars with blended images and finite disk size (not colorless), and so forth. Luckily all the raw data are being archived, and so it has proven possible to go back and re-evaluate all the observations with the new templates. It nevertheless remains quite difficult for the observers to evaluate precisely what their level of completeness is, and the methodology can be described as that of successive approximations. At least three other similar searches are underway; they are acronymed OGLE (Optical Gravitational Lens Experiment), EROS, and DUO.

Astone's poster (Ch. 22) dealt with an on-going search for gravitational radiation bursts using bar-type antennas located in Europe. Schutz's talk (Ch. 13) discussed the somewhat similar problems of the planned search using interferometric detectors to be located both in the US (LIGO) and in Europe (VIRGO and perhaps others). Here the problem is that the template must come from theoretical calculations of what you expect from the merger of a pair of neutron stars or the collapsing core of a supernova in terms of time-varying quadrupole potentials. Numerical general relativity is not yet fully equal to the task, and the theoretical part of the project is currently in the "method of recurrent simulations" stage.

Curiously, many participants were left with the impression that LIGO and VIRGO would constitute the first deliberate searches for gravitational radiation. In fact, J. Weber has operated two or more bar-type antennas continuously for more than 30 years, and he first described the design at a meeting on general relativity and gravitation in Warsaw in 1962. Peter Bergmann, who worked with Einstein, predicted at the time that it would be a century before anything came of the effort. He is, so far, 1/3 right.

A related problem arises when you are carrying out one of the classification projects mentioned in §3.2 or a corresponding program in the time domain and discover that none of your templates is a good fit to a particular object or image or light curve. The poster of Arenou et al. (Ch. 46) concerned one of these cases. They have been involved in the primary (astrometric) part of the HIPPARCHOS mission and discovered that a few of the objects in the catalog acted neither like single stars nor like double ones. They had in fact (re)discovered what are called astrometric binaries, ones where you see only a single image, but the orbital image makes it wiggle around in the sky over a few years.

21.3.8 *One population or two?*

"Alan of Lille (de L'Isle): Alan of Tewksbury? Alan of Tewksbury: Alan of Lille?" Robert K. Merton 1965

An astronomer quite often finds himself holding two handfuls of beans and wanting to know whether they came out of the same bean bag. Exam-

ples include active galaxies selected at different wavelengths (radio, optical, X-ray); the Lyman α forest of absorption at low *vs.* high redshift; the planetary nebulae in different types of galaxies; the chemical compositions of meteorites *vs.* those of asteroids; quasars with and without broad absorption lines; and many others that would take more words to explain. Problems of this sort were not discussed, apart from Rood's poster, which attempted to decide whether gamma ray bursts and Abell clusters are drawn from the same parent population (his answer is, partly, probably).

Sometimes one population is a real set of observed sources (etc.) and the other a theoretical model that predicts a range of properties. In this case, the customary astronomical approach is the Kolmogorov-Smirnov test, which is, however, not necessarily the best available.

Deciding whether an apparent cluster of data points can or cannot have arisen by chance out of a smoother underlying population is a somewhat similar problem. Efron's example (childhood cancers in California towns) was not astronomical, and Berger's (Ch. 2) "bivariate observations" came from an astronomical, but otherwise undefined, context. We were assured that the former cluster was not significant and that the latter ones were.

21.3.9 *Censored and truncated data: The problem of sample selection*

"Most mistakes are made before ever putting pencil to paper." VT (frequently)

Which is which, for starters? Censored data occur when patients are lost to follow-up (but you have no reason to support *a priori* that they are different from the ones you can discuss). Truncated data occur because patients enter a trial after the first year, and so you have only 3 or 4 or 5 years of follow-up, not the full six years of the study (and, having entered later than the others, they might be somehow systematically different). To take an astronomical example, if you search for X-ray quasars by staring with an optical sample, then your data on non-detected ones will be censored — you will know that the quasars exist and that they have fluxes less than your detection limit, but you will not know the actual flux values (and have no reason to suppose that they are systematically different from the nearby sources whose fluxes you can measure). An X-ray search will lead to a truncated sample — below your flux limit, you will not only have no numbers, but you will also not be sure that any such objects exist.

One must be clear at the outset that there are no foolproof ways of dealing with these situations, except to work harder and find the missing or undetected objects (and then, of course, you will try to analyze the new, larger set, and be right back where you start from, except with a lower flux limit). Thus discussions of statistical methods can address only which approaches are likely to be best for a given, physically-defined, situ-

ation, and, very importantly, which estimators are equivalent to or reduce to others in specific situations. Yang (Ch. 5) provided a quite-technical review of some ways of dealing with such incomplete samples, including the Kaplan-Meier (two-people) survival curve for right censored data and the Lynden-Bell (one person) estimator for non- parametric analysis of incomplete data. Apparently all methods that bin data points reduce to the Lynden-Bell method where there is only one point per bin. Caditz's poster (Ch. 37) considered a smoothing (*vs.* binning) non- parametric approach to estimating luminosity functions from truncated (flux-limited) samples. The topic was also reviewed by V. Petrosian in SCMA I.

Many real astronomical samples are bounded in such complex ways that knowing how to cope with "simple" truncation or censoring is not very much help. In these cases, what we really need is guidance on how to choose the samples of stars, galaxies, or whatever to look at. The brute-force approach of gathering enormous quantities of data and throwing out all but the many- dimensional corner of completeness is essentially never possible.

A typical project (one I have been involved in for 20 years or so) is attempting to determine the distribution of mass ratios of binary stars (as a guide to understanding the processes of star formation). Recognizing that a star is a binary and getting enough information about it to know the mass ratio is dependent on apparent brightness (that is, real brightness/mass and distance), the period of the binary (with two or three peaks of high detectability), the mass ratio itself, the eccentricity of the orbit, the evolutionary stages of the two stars, and probably other things; and there is every reason to expect the mass ratio distribution to be a function of at least some of these. The result is that people who have examined different subsets of archival data and/or added to it get wildly different answers and feel so strongly about them that we can only just barely be brought to acknowledge that it is a problem in sample selection (and no samples are perfect), and not one in abnormal psychology.

21.3.10 Data compression and storage

"Archiving is easy; retrieval is difficult." Speaker at a 1993 IAU Symposium on Wide Field Imaging

Actually, with the advent of terabyte projects like the Sloan Digital Sky Survey, MACHO searches, 2MASS, and so forth, even archiving isn't all that easy, resulting in the need for very efficient ways of compressing data with little or no loss of information. Murtagh and Scargle (Ch. 7 & 19) in their talks both mentioned that wavelets show great promise for efficient compression and storage of many different kinds of astronomical data.

21.4 Powerful methods

"In my head are many facts that, as a student. I have studied to procure. In my head are many facts of which I wish I was more certain I was sure."
Oscar Hammerstein II (The King and I)

Historically, astronomers have found their correlations, templates, clusters, and all the rest by methods that could not be easily quantified or even taught, and turned to quantifiable, statistical ideas only afterward to answer the question "how sure am I that this is right?" For instance, only Fritz Zwicky could find Zwicky clusters on Palomar sky survey plates, but the catalog remains useful down to the present, and we can now say a good deal about how complete it is for different kinds of clusters at different distances from us. Kolmogorov-Smirnov, chi-squared, and other tests are meant to answer that question. In contrast, least-squares, by the time you push the last button on your calculator, has provided not only the coefficients you asked for but also some kind of error estimate. Some, but not all, of the more elegant techniques presented at the symposium simultaneously fit something to something and evaluate the goodness of fit.

21.4.1 *Bayesian (vs. frequentist) statistical methods*

"At some level we are all Bayesians" Michael S. Turner at a 1996 symposium on cosmic distance scales. *"Yes. Those who are not were run over by trucks at an early age."* VT, same symposium. By this was meant that we all carry around with us a great many prejudices about the relative likelihood of various outcomes to everyday experiments, like crossing the street during the rush hour, and that we would be worse off without them. The other side of this coin is that it is possible to come to believe something so firmly that no amount of later data can influence one's opinion; on both scientific and non-scientific issues.

My own impression (prejudice?) is that Bayesian methods are most useful when one expects to change one's mind only slightly. Jeffery's talk (Ch. 3) on use of lunar laser ranging data to improve our knowledge of lunar motions and earth rotation is a good example. In contrast, I do not think there has ever been a time (including now) when applying such methods would have improved our knowledge of the Hubble constant, for the problems have always been wildly wrong choice of *a priori* probabilities, neglect of important physics, and inappropriate data selection, and I suspect they still are. Religious conversion is probably also not amenable to Bayesian treatment (perhaps it is a critical point phenomenon?).

Berger's review (Ch. 2) of Bayesian analysis was, on the whole, comforting to an outsider, in the sense that it left impressions (a) that exact choice of prior probabilities makes relatively little difference if you have a reasonably large data set, (b) that modern computational methods make it relatively easy to get hold of the terms that enter into the posterior proba-

bilities, and (c) that a Bayesian answer will differ wildly from a frequentist one about how probable a particular hypothesis/event/etc. is only under rather pathological circumstances. Pursuant to (a) and (b), he present a default option for choosing the prior distributions for the unknown parameters of each model, which can be used when you don't quite know what else to do. One of the surprises was that there are conceivable data sets that are not very probable under any hypothesis, and I am not sure whether the point being made was much more profound than that it is "very unlikely" that I shall see a license plate number AAA 222 on the street today, but no more unlikely than PZD 181.

Bayesian analyses of specific problems were presented in the posters of Núñez & Llacer (Ch. 28) (image reconstruction) and Kester (deconvolving camera and source contributions to spectra from the Infrared Space Observatory).

21.4.2 *Wavelets and related transforms*

"Mathematics may be compared to a mill of exquisite workmanship, which grinds you stuff of any degree of fineness; but, nevertheless, what you get out depends on what you put in; and as the grandest mill in the world will not extract wheat-flour from peascod [peapods], so pages of formulae will not get a definite result out of loose data." Thomas Huxley, 1869

And wavelets, the topic most discussed at the symposium, can grind exceedingly small (but, of course, still subject to the Huxley limit that, in modern language, is generally described as GIGO: garbage in, garbage out). The prominence is apparently a recent development. Wavelets had only a single index entry in SCMA I and no definition in the glossary. Thus I provide my own crude one: wavelet transforms are a lot like Fourier transforms, but you get more choices. In particular, the shapes can be chosen to be especially good at localizing edges, bumps, and discontinuities (at which sine and cosine waves are rather poor) and to pick out several widely spaced length or time scales where there is lots of information, without having to worry too much about the ones in between.

One or both of these virtues was conspicuous in most of the poster presentations where wavelets were applied to specific kinds of astronomical data. These included Lawrence et al. (temporal structure of solar activity, Ch. 35), Lehto (light curve of OJ 287, Ch. 36), Anderson & Langer (image processing), Young et al. (gamma ray burst time series, Ch. 37), Chereul et al. (the local gravitational potential determined from HIPPARCHOS data, Ch. 32), Damiani et al. (source detection with photon counting detectors, Ch. 33), Kashyap (flat fielding and source identification, Ch. 34), and Bijaoui (multiscale analysis, Ch. 10).

The speakers addressed both general properties and specific applications of wavelets. Several of them mentioned that the "Mexican hat" is quite commonly used, but not entirely proper for most applications; Haar wavelets,

in contrast, are like falling off a log. Murtagh (Ch. 7) emphasized the problems of data selection and coding in large data bases, using a number of IAU spectra of the Seyfert galaxy NGC 4151 as an example of how principal component analysis can be achieved economically. Orthogonal and discrete wavelet transforms are particularly suited to some applications.

Scargle reviewed (Ch. 19) both methods and existing astronomical applications. These range from initial data processing through image and time series analysis to data compression, transmission, and storage. For instance (later than the time frame Scargle reviewed) the Mars96 camera data will be sent back in wavelet transform. Hazards abound: scalegrams (analogous to power spectra in Fourier transforms) are not the same as scalograms (the absolute values of the wavelet coefficients), and “matching pursuit” is not at all what you might think. But his main point was that wavelets provide a whole new way of thinking about astronomical (etc.) data, and not must a library of specialized techniques.

Priestley (Ch. 16) addressed wavelet analysis for the study of time-dependent spectra and presented a number of caveats about things that wavelets cannot do, starting with the analog of the Heisenberg uncertainty principle, that you buy high resolution in the time domain at the expense of resolution in the frequency domain, and conversely. Bijaoui (Ch. 10) focussed on analysis of images that have interesting structures on many different scales, using, for instance, wavelets to pick out a spiral arm in a poorly-resolved galaxy image.

Self-respecting wavelets come in complete sets. There are also interesting transforms with respect to incomplete sets of basis functions, for instance the Gabor transforms mentioned by Scargle and used by Schaefer & Bond in their search for quasi-periodic oscillations in AM Her stars.

21.4.3 Neural networks

“If I see a sheep on a hill, I think there is likely to be a wolf nearby.” Leon Sibul (during panel discussion)

Artificial neural networks are programs/algorithms that are supposed to work somewhat the same what the human brain does. That is, one has a bunch of digital nodes (neurons) connected by input/output channels (axons and dendrites) through a bunch of synapses, and what goes out the axons depends in some complex way on what came in the dendrites. Artificial neural networks have in common with humans that they can, in some sense, learn from experience and get better at a task after doing it for a while. Eventually, perhaps, we should expect boredom to set in and then, as the networks become more human, they ought to start generating totally unexpected outputs, as, for instance, the wolf in a discussion of what you can infer about the colors of all sheep on the basis of seeing one that is black on at least one side.

Among the poster presentations, Nair & Principe applied ANNs to long and short time scale prediction, Naim (Ch. 44) to identifying peculiar galaxies, and Bailer-Jones (Ch. 42) to classification of stellar spectra. I believe that all poster contributors and invited speakers has applied biological neural networks (of the highest quality) to their presentations.

21.4.4 *Other methods*

“You can trust us ...” James Berger (at coffee break). If only because there are so many ways to do things, one of them must give the right answer.

Statistical techniques that were neither obviously Bayeslet or wavesian appeared in a number of posters. These included multifractals (Lawrence et al. on solar activity, Ch. 35), multiscale maximum entropy (Starck & Pantin on image reconstruction, Ch. 29), minimum distance indicators (Qian on autoregressive models), genetic programs (Kester on the ISO camera), maximum likelihood (Luri et al. on statistical parallaxes from HIPPARCHOS data, Ch. 39), bootstraps and other Monte Carlo methods (Hertz on X-ray binary periods, Ch. 23, and Hesterberg on error estimation), two-point correlation functions and nearest neighbors (Brainerd on the existence of gamma-ray repeaters), and probably some others. I am not claiming with high confidence that all of these are truly distinct, only that they have different names.

21.4.5 *Comparisons and relationships of methods*

“How to lie with statistics” Title of a 1960’s book. *Statistics and Truth* Title of a 1995 book (by C. R. Rao)

A particularly interesting aspect of the talks by Efron and Martinez (Ch. 9) was their comparisons between methods and applications from different universes of discourse. Efron alternated between biomedical and astronomical problems, showing, for instance, that survival curves are a good deal like $\log N - \log S$, and that hazard rates, Poisson regression, nearest neighbors, and Mantel-Haenszel or log-rank tests can be useful concepts in both.

Martinez attempted to identify the standard statistical methods that correspond most closely to a variety of techniques that astronomers seem to have invented for themselves. For instance, “our” two-point correlation function, $\xi(r)$, is quite similar to “your” second order intensity function. The speaker did not mention the histories of the various methods, so that no one could be tempted to say about anyone else, “Ah yes; he has re-invented the wheel. Only his is square.”

21.4.6 *Plausibility tests*

"Ask a silly question, you get a silly answer." Virginia Farmer Trimble (author's mother) c. 1953

That one's results should make sense is obvious, though only about four presentations specifically mentioned the point (posters by Starck & Pantin, Valdes, and Hertz, which last revived two significance tests from the pre-war astronomical literature, and Sibul during one of the panel discussions, Ch. 29, 23). It is a very old problem that pops up again and again. For instance, the first spectrogram of quasar 3C 273, taken in 1963, had a handful of emission lines whose wavelengths were fit quite reasonably well by high ionization states of neon and such. The fit to part of the Balmer series of hydrogen was statistically no better, but, since hydrogen is the commonest element in the universe, there was really no competition.

Human beings make such judgments easily, quickly, and naturally (and quite often wrongly). Building physical sense into processing algorithms seems to be rather difficult, and many published astronomical papers mention a "biomechanical servo" stage, in which things that sound silly are thrown out of the sample. Ancient examples include pixels with negative fluxes, supernova candidates nowhere near any galaxy, spectra with no features, and binary systems in which the star you don't see is more massive, and so should be brighter than the one you do see. This is, of course, a good way to fail to discover new classes of objects. And no, I don't know what the solution is, except that a group of systems of the fourth type were once a set of black hole candidates called Trimble-Thorne systems. None actually has a black hole, but they all turned out to be interesting in terms of stellar physics.

21.4.7 *Implementation and looking ahead*

"The future is not yet in existence." Edward Wegman

Most day-to-day analysis, from image processing on up to simulations of interstellar clouds and large scale structure of the universe, is done with standard software packages (and a very good thing, too, cf. the square wheels of §4.5). Wegman (Ch. 11) described a number of these packages, what they do, and how they can be grabbed out of the ether. Not everything he mentioned is younger than yesterday's newspaper. Though the websites to which he provided points all date from the last couple of years, his paper references go back to 1953.

The scientific organizing committee left the participants with several "thought problems" in the area of what should be done next. Would an actual, paper, book of modern statistical recipes be useful? What about a web site, and who would maintain it? [*A web site with on-line statistical software for astronomy has now been initiated at <http://www.astro.psu.edu/statcodes>. Eds.] How can we foster collaborations between statisticians and astronomers*

that will be attractive to both in the sense of advancing basic knowledge in both fields so that each collaborator has something significant to add to his CV at the end? Participants nodded solemnly and promised to Think About it All.

Acknowledgments: It is the traditional prerogative of the last speaker at a conference to extend the collective thanks of the participants to all those who have made the gathering possible. Our sincere gratitude goes to Dean Geoffroy and Pennsylvania State University for their hospitality; to the seven sources of silver who provided financial support, NASA, NSF, ISF, IAU (but not much), Eberly College, ISI, and IMS; to Drs. Babu and Feigelson for putting all the science together; and to Debbie Noyes (and undoubtedly others) for help with logistic arrangements. We all look forward to gathering again at SCMA III in about 2001.

REFERENCES

- [1] Lutz, T. E. and Kelker, D. H., 1973, *Publ. Astron. Soc. Pacific* 85, 573.
- [2] Malmquist, G., 1920, *Lund (Observatory) Medd. Ser.* 2, No. 22, 32.
- [3] Ryle, M., 1955, *Observatory*, 75, 137.
- [4] Scheuer, P. A. H., 1953, *Proc. Cambridge Phil. Soc.* 53, 764.
- [5] Schmidt, M., 1968, *Astrophys. J.* 151, 393.
- [6] Scott, E. L., 1956, *Astron. J.* 61, 190.