A View of PHYSTAT 2003

John A. Rice
University of California, Berkeley, 94720 USA

1. INTRODUCTION

I present the view of a statistician for whom the landscape of statistical methods in high energy physics (HEP) and astrophysics was initially very foggy. As the fascinating conference progressed, the fog partially cleared and some of the underlying issues began coming into clearer focus. Because of my perspective, this summary is by no means comprehensive, but rather selectively concentrates on some of the statistical issues that arose during the conference. I thus apologize to the many participants whose contributions are not adequately reflected in what follows.

2. METHODOLOGICAL OVERVIEWS

The conference featured a number of valuable overview talks, some focused on statistical methodology and some on the underlying scientific issues.

Physicists might find it helpful to view the contrasting perspectives of Diaconis, Friedman, and Stark as spanning a large portion of the space of current statistical thinking: contemporary high-dimensional Bayesian methodology, pragmatic empiricism, and a hard-line frequentist view.

Many other talks can be placed within this barycentric coordinate system. Efron's contained substantial projections on all three, illustrating how the interaction between classical maximum likelihood analysis, Stein estimation, and Bayesian modeling of high dimensional Gaussians has enriched our empirical understanding of estimation in this prototypical context. The hierarchical models and smoothing techniques presented by van Dyk illustrated the modern Bayesian approach at work on an interesting concrete problem in astronomy. Reid provided a valuable, concise presentation of classical and recent frequentist methods of dealing with nuisance parameters. Genovese illustrated the use of modern methods of smoothing and simultaneous inference for curves, power spectra in particular.

Of the scientific overviews of statistical methodology, those of Nichol and Prosper echoed the empiricism espoused by Friedman. Barlow's presentation of statistical issues in HEP was much more classical and less broadly inclusive than that of Feigelson on statistical issues in astronomy. This contrast may be in large part due to fundamental differences between the ways data are collected in the two areas, HEP relying on carefully designed experiments, and astronomy being necessarily observational. Formal inference thus plays a larger role in the former and exploratory data analysis a larger role in the latter. Similarly, formal methods of inference dominated James' presentation. This contrast persisted in the presentations of Porter and Digel, which provided for the statisticians a concrete sense of the scope of statistical methods in the two sciences.

3. SOME SPECIFICS

In addition to the general overviews, I gained some particular insights relating to the use of statistical methods, among them:

What is meant by $180.1 \pm 3.6 \pm 4.0$? Physicists are deeply concerned with systematic bias, much more so than is usual in statistics (Sinervo gave a very instructive overview). Variance is small in these high precision measurements, whereas in more typical statistical applications, bias is often swamped by variance.

The null hypothesis is often seriously entertained in HEP: the hypothesized phenomena may not exist. In many statistical applications, the null hypothesis serves much more the role of a strawman.

HEP experiments are done with extraordinary care—witness the clever techniques of blinding (Roodman).

Small counts are ubiquitous in HEP, and are frequently as Poisson as one can imagine for actual data.

As a consequence of the previous point, classical results in theoretical statistics, such as Neyman similar tests, are really of substantial importance, whereas to many statisticians these appear as the artifacts of a distant past. It is gratifying that some scientists really care about uniformly most powerful unbiased tests!

In addition to these classical concerns, more contemporary developments in statistics and machine learning play an important role, not only in astronomy, but also in HEP. Examples included the presentations of Askew, Cardossa, Cranmer, and Knuteson among others.

Care in constructing probability models, commensurate to that used in carrying out the experiments, can really pay off (Canelli).

Physicists rely on Monte Carlo (MC) to a degree surprising even for statisticians. (In fact, many MC methods have their roots in physics). For example, goodness of fit tests are often performed relative to a distribution that can only be expressed via Monte Carlo.

Finally, despite the best efforts of several patient physicists, "cuts" remain a mystery to me, but an interesting one.

4. FORMAL METHODS OF INFERENCE

Both Bayes and frequentist methods of formal inference rely on mathematical idealizations, such as probability spaces, measurable functions, independence, and distributions which are exactly known (at least up to parameters). HEP seems as well suited to this imaginary world as any scientific field—counts as Bernoulli as one could hope, experiments done with extraordinary care, careful and detailed modeling, both parametric and via MC, and precise numerics (Quayle). Using these concepts correctly requires care in modeling and in getting the math right (Linneman, Demortier).

Formal methods being applicable, there still remains the question of how to relate them to the science. This theme was present, at least implicitly in many of the talks and explicitly in others, for example Porter, Punzi, Rolke, Shawan. It is interesting that the HEP community seems to have settled on requiring more stringent standards for discovery than for setting upper limits when using the frequentist paradigm. There is also the use of code words, "evidence for" and "discovery of," corresponding to different significance levels. A detailed explication of why the 5σ level has been apparently accepted as conclusive evidence would be quite interesting.

Relating formal Bayesian methods to the science is not straightforward either, as illustrated by Shawan. Savage [3] wrote that one cannot enjoy the Bayesian omelet without breaking the eggs, but even having agreed in principle to make an omelet, many further decisions have to be made. Whether one enjoys the omelet depends up the quality of the chef, the quality of the ingredients and the taste of the diner. HEP and astronomy certainly offer high class ingredients and we have seen several skillful chefs at work, for example Diaconis, Loredo, Scargle, and van Dyk. One must choose what style of omelet: subjective omelet (Savage), uninformative omelet (Jeffreys), hierarchical omelet (van Dyk), least favorable omelet (Stark), or the truly exotic omelets concocted for high or infinite dimensional parameters. One wonders how these decisions could be made collectively by the army of chefs of an HEP collaboration.

In the statistics community, the Bayes/Frequentist wars have waned, in part because many of the old combatants have perished, and in part for reasons I discuss in the next section. The theme that contemporary statistics is not prescriptive emerged repeatedly: "much depends on attitudes and psychology" (Efron); "Bayes and frequentist are not opposing points of view" (van Dyk); James teaches both in parallel. The interpretation of statistical results within a scientific context has a large subjective component, and it is implausible that there exists a single, formal paradigm for how one learns from data, even setting aside the key role of informal exploratory data analysis.

5. STATISTICS IN HIGH DIMENSIONS

High dimensional statistical phenomena were a feature of many talks, and effective treatment of these sorts of problems is hard whatever point of view one takes, Bayes, frequentist, or opportunistic empiricism.

Maximum likelihood estimates behave badly in high dimensions, fluctuating too wildly as Efron pointed out for the canonical multivariate normal case. The unfolding problem is another well known example. A common method of stabilizing maximum likelihood estimates is via penalization, so that rather than maximizing log-likelihood(θ), one maximizes log-likelihood(θ) + J(θ,λ).

J(θ,λ) can often be usefully interpreted as the log of a prior on θ, or may arise directly from a prior, bringing Bayes and frequentist perspectives into close alignment. (Scargle's presentation is one example). But life is not so simple as just specifying a prior and integrating—the choice of prior really matters, as again illustrated by Efron. While we may think we may understand what it means to put a uniform prior on a scalar parameter belonging to the unit interval, what it means to put priors on high dimensional parameters is very hard to understand. Our experience of the relationship of geometry and measure in three dimensions does not provide good intuition for high dimensional Euclidean space, where:

- Almost all the volume of the unit cube lies near the corners.
- Most of the mass of the cube lies close to any subspace of R^n.
- Almost all the volume of the Euclidean ball lies near the edge.
- Gaussian measure concentrates on a thin shell far from the origin.

(For these and other cheerful facts, see [1]).

Of course the prior on the unit interval can "matter" in one dimension as well, which is reasonable from the classical view of a prior as personal probability. However our lack of intuition about measure and geometry in high dimensions and the exotic priors (cf. Diaconis, van Dyk, Scargle) makes interpretation in terms of personal belief at best highly formal. A 95% posterior region resulting from such priors is not 95% "credible" in the ordinary English language sense of the word. It is thus imperative to assess such procedures in terms of their frequentist and empirical properties, which again brings the perspectives into closer alignment.

We have heard relatively little about empirical Bayes procedures in this meeting, but the times seem ripe. In astronomy, for example, one measures repeatedly the properties of many, many similar objects sampled from a universe of such objects. It would thus make sense to not treat the analysis of each measurement de novo, but to integrate over the accumulation of experience. However, this is easier to say than do, and there are real challenges to be met in developing such methodology.

6. BEYOND FORMALISMS

There is more to statistics these days than dreamt of in the philosophies of Bayes, Fisher and Neyman. As I. J. Good wrote [2], "In our theories we rightly search for unification, but real life is both complicated and short, and we make no mockery of honest ad-hockery." Computing power has and continues to produce a rich buffet of procedures, as illustrated in the presentations of Knuteson, Prosper, Friedman, Nichols, Vilalta, Lee, Cardosa, Askew, Cranmer, and Levi.

Some may be concerned that this explosion of computing power coupled with unfettered imagination is leading to the degeneration of the stern discipline of statistics established by our forebears to a highly esoteric form of performance art. Looking over the spectrum of current work in statistics, one in fact does see far too many examples consisting of a proposed algorithm, a small number of simulations over a restricted range of scenarios, and maybe one real example.

There are, thankfully, constraints on unfettered imagination. First, one can try to do the math right; asymptotic analyses and considerations of optimality can provide insight into finite sample behavior. Extensive simulations, informed by such mathematical insight as is available, are valuable in sorting the wheat from the chaff. And finally there is the crucial test of utility with real data in real scientific contexts, revealing whether proposed methods actually lead to sensible and meaningful results. There is hope that evolutionary pressures will lead to extinction of the frail and survival of the fittest.

References

[1] K. Ball, "An Elementary Introduction to Convex Geometry", Mathematical Sciences Research Institute, 1977.

[2] I. J. Good. *Estimation of Probabilities.* MIT Press, 1965.

[3] L. J. Savage. *The Foundations of Statistics.* Wiley, 1954.