

“Numbers and reasons” — core elements of numerical inference techniques today, with implications for the rhetoric of LASR presentations

Fred L. Bookstein

University of Vienna and University of Washington

If the overall rubric of this year’s LASR is “high-throughput sequencing, proteins, and statistics,” then this talk falls squarely under the “and statistics” prong. Yet as some of you already know, I have long been dubious about that word “and” in our banner. I fear that over the course of our excursions into bioinformatics the essence of reasoning from quantifications, as it coalesced over a dozen decades of progress in the natural sciences, has seriously deteriorated. Our field has drifted steadily farther and farther from its intellectual roots in the rhetoric of the sciences of natural explanation, roots too infrequently acknowledged.

Of late this unease of mine has been one stimulus driving the drafting of a textbook for the undergraduate course “Numbers and Reasons” that I have been teaching for a while, first at the University of Michigan, later at the University of Vienna, and now at the University of Washington. To buttress the book I have assembled a range of dicta and a roster of good canonical examples of inference from numerical evidence across the natural and social sciences. There is a surprising disharmony between the language and rhetoric of these indubitably useful inferences and the language and rhetoric that we generally teach our students to emulate. This talk, sampling from the themes of the book manuscript, sketches some of the main themes of numerical inference as they apply in its most successful and widely disseminated examples. My intention is to provoke you into a reconsideration of what statistical forms of reasoning have to offer the intellectual natural scientist here in 2010, even as bioinformatic or proteomic pattern analyses per se seem to go forward better *in silico*.

The central conceit upon which the textbook’s argument turns is the iteration of several dogmas. Here I introduce them all at once, *en bloc*:

Dogma 1. *The ITT.* “It has been called the interocular traumatic test; you know what the data mean when the conclusion hits you between the eyes.” — W. Edwards, H. Lindman, & L. J. Savage, 1963.

Dogma 2. *Abduction.* “The first starting of a hypothesis and the entertaining of it, whether as a simple interrogation or with any degree of confidence, is an inferential step which I propose to call *abduction*. I call all such inference by the peculiar name, *abduction*, because its legitimacy depends upon altogether different principles from those of other kinds of inference. The form of inference is this: The surprising fact, C, is observed; but if A were true, C would be a matter of course; hence, there is reason to suspect that A is true.” —C. S. Peirce, 1903.

Dogma 3. *Passion.* “Science, no less than theology or philosophy, is the field for personal influence, for the creation of enthusiasm, and for the establishment of ideals of self-discipline and self-development. No man becomes great in science from the mere force of intellect, unguided and unaccompanied by what really amounts to moral force. Behind the intellectual capacity there is the devotion to truth, the deep sympathy with nature, and the determination to sacrifice all minor matters to one great end.” —Karl Pearson, 1906.

Dogma 4. *Science is cognitive, not empirical.* “The route from theory or law to measurement can almost never be travelled backwards. Numbers gathered without some knowledge of the regularity to be expected almost never speak for themselves. Almost certainly they remain

just numbers.” —Thomas Kuhn, 1961.

Dogma 5. Unreasonable effectiveness. “The enormous usefulness of mathematics in the natural sciences is something bordering on the mysterious, and there is no rational explanation for it. ... It is not at all natural that ‘laws of nature’ exist, much less that man is able to discover them. ... The miracle of the appropriateness of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve.” —Eugene Wigner, 1960.

Dogma 6. Consilience. “Trust in consilience is the foundation of the natural sciences.” — E. O. Wilson, 1998. [Consilience means the agreement of inferences arising from different sets of systematic observations. In a numerical context, it becomes the equality of estimates of the same quantity from different experiments.]

These principles can often be woven directly into the narratives we use to convince students of the importance of numbers and measurements throughout the modern sciences. The Fourth Assessment Report (2007) of the Intergovernmental Panel on Climate Change (IPCC), for instance, rests its main assertion about anthropogenicity of observed global warming upon an interoptical trauma test involving one figure, which appears as Figure TS.23 on page 62 of the 2007 publication and also as Figure 9.5 on page 684 (where it is accompanied by a great deal more detail). Certainly it “hits you between the eyes” (Dogma 1) that the outputs of the models with anthropogenic inputs separate cleanly from the outputs of models without those inputs over the last few decades of data and (Dogma 2) that the measurements align only with the former. The trust in models, in turn, illustrates dogmas 5 and 6, the fundamental trustworthiness of mathematical models in the natural sciences and the expectation that measurements at one level (say, temperature at a point) must be consistent with the partial differential equations relating such measurements over the entire planet. The objections to this report, if you will let me overgeneralize for a moment, arise mainly in connection with Dogma 3 — the “passion” for explanation is easily twisted into an accusation of bias by political opponents of the policy implications of findings like these — while Dogma 4 is implicit in the reliance on the differential equations, the paths to explanation, rather than the raw temperatures, the “mere measurements.” You cannot forecast global warming adequately by extrapolation; you must proceed via the equations, which is to say, understanding.

Turning to concerns more directly apposite to this conference of ours, let us closely inspect the founding document of all of bioinformatics, the 1953 paper by James Watson and Francis Crick announcing the “double helix.” In a paper barely one page long, about 900 words, as expanded to book length by Watson in his best-selling memoir *The Double Helix* of 1968, we note all of the following: (1) the essence of the discovery in an instant, the shape matching of the two versions of paired bases, (2) the further matching to direct measurements from Rosalind Franklin’s diffraction image, (3) the fit of Watson’s physical model to the atomic spacing standards in Linus Pauling’s textbook, and (4) the rationalization of the previously mysterious “Chargaff’s rules” for the equality of the counts of A and T and of G and C. None of these inferences appear to be “statistical” at all. The statistics are hidden in the physics of diffraction images and their photography, atomic spacing data and their uncertainties, and the like.

How do our dogmas apply here? The first, the ITT, is explicit in the standard diagram of the matching shape of the A-T and G-C base pairs. Watson says that he “suddenly became aware” of this, whereupon the whole scheme fell into place immediately. The second, abduction, appears twice, once accounting for Chargaff’s rules and a second time in the successful modeling of the measureable parameters of Franklin’s image. The third, passion, runs like a leitmotif through Watson’s recounting of his time with Crick — back in that golden age, it was the desire to get at the truth that drove biologists like these two. Dogma 4, on the limited role of raw em-

piricism, was illustrated inadvertently by Chargaff's own work: he reported that the counts of A and T were approximately equal when, in fact, they must be *exactly* equal. Chargaff so badly misunderstood the salience of his own measurements that to his death (only a few years ago) he insisted he had been cheated out of his part of that Nobel Prize. Dogma 5, on the effectiveness of mathematics, underlies the interpretation of instrumental data at every step: the mathematics of the Fourier domain accounts for the diffraction images, for example, and the simple mathematics of multiplication allowed Crick to check atomic spacings on the model with the aid of a slide rule. Finally, Dogma 6 (consilience) is ubiquitous: measurements of the same system at different levels of analysis *must* result in the same numerical inferences. The DNA molecule must obey the laws of physics, of solutes, of diffraction, and of information theory, all at the same time; there is no *élan vital*.

Still in the domain of molecules (and likewise of Nobel Prizes) we can inspect Jean Perrin's demonstration that atoms exist. Perrin's book *Les Atomes* of 1911 (English, *Atoms*, 1923, still in print), was one great abduction based on the approximate equality of fourteen fundamentally different experimental estimates of Avogadro's number (the number of hydrogen atoms in one gram of the gas). In a hundred pages of précis of exquisite experiments involving microscopes, telescopes, and a huge range of detectors in-between, Perrin showed (1) that Einstein's model of Brownian motion fits both translational and rotational modes for "molecules" that are large enough to see as individuals in an image, (2) that the Maxwell-Boltzmann distribution (velocities distributed as a spherical Gaussian) is approximately correct as a model for those displacements, and (3) that either of these models is consistent with the partial-pressure model that had accounted for rarefaction of the atmosphere with height for more than 100 years. (As the Gaussian distribution arises ultimately from Stirling's approximation to the factorial function, furthermore, one can claim that all of these findings rest ultimately on large-scale properties of the integers.) All of the dogmas apply to deconstructing Perrin's work just as to the IPCC's or to Watson and Crick's. Regarding Dogma 3, "passion," for instance, Perrin writes, in the reserved tone of the stereotypical natural scientist, that "it was with the liveliest emotion" that he checked his estimate of Avogadro's number from the microscope with the one derived from measurements of viscosity, and, much later, that "our wonder is aroused" at the agreement of all fourteen of these variegated experiment-based calculations. Only his hypothesis, the existence of atoms, survives the challenge of all these data (Dogma 2, abduction). The mathematical foundations of all this (Dogma 5), from the Gaussian distribution itself on forward, are pervasive, as is the fundamental trust that measurements of a little ball visible in a microscope must be consilient (Dogma 6) with the quantitative properties of atoms a hundred million times smaller than that. Perrin's demonstration, one of the great triumphs of statistical science, deserves to be far more widely appreciated today than it actually is.

The standard theme of statistical inference, at least in its vulgar sense of null-hypothesis testing, is startlingly absent from these stories. Kuhn puts it well in a wisecrack early in his 1961 article about the role of measurement in the physical sciences when he notes that in the textbooks, no theory is considered to be tested by any test that it has not already passed. In textbooks the quantitative hypothesis under study is factually true: it fits the data. The true theory becomes the null hypothesis, and so, naturally, it stands unrefuted: the significance test turns into a goodness-of-fit test instead, to be carried out in a particularly demanding way, via explicit inspection of noise distributions in a skeptical frame of mind. In place of p -values there is an unusual concentration on crucial details of measurement — where suites of variables come from, how calibration of machines can maximize their reproducibility — that are almost always overlooked in textbooks of statistical method. The exceptions to this generalization, such as the 1989 book by Harald Martens and Tormod Næs on multivariate analysis of mass spectrograms or (I note modestly) my 1991 book on the biomathematical foundations of morphometrics, are

rare but, when successful, prove to be citation classics partly by reason of that rarity.

But they all share one central rhetorical concern: consilience (Dogma 6), the convergence of evidence from multiple sources. Now consilience requires a relatively deep understanding of the way that such multiple sources relate to a common hypothesis. To have a reasonable chance of making sense in these domains we must take real (physical, biophysical) models of system behavior (the organism on its own, and the organism in interaction with our instruments) as seriously as we typically take abstract (statistical) models of noise or empirical covariance structure. Serious frustrations and paradoxes can easily arise in this connection. Over in the psychological sciences, Paul Meehl once wisecracked that most pairs of variables are correlated at the so-called “crud factor” level of ± 0.25 or so. It is this correlation, not a correlation of zero, that represents an appropriate null hypothesis in these sciences. Closer to home, in my own application domain of morphometrics, landmark shape distributions are never spherical in shape space. The broken symmetries are properly taken not as algebraic defects in our formulas, but as biological aspects of the real world; they are signal rather than lack of fit. A few years ago Kanti Mardia, John Kent and I published a quite different model for a-priori ignorance in morphometrics, an intrinsic random field model in which noise is self-similar at every scale. I am still awaiting news that somebody has tried to fit their data to *that*.

That word “consilience” was originally William Whewell’s. He introduced it in 1840 by reference to the way that approaches to the same scientific nexus by different instruments or principles often resulted in consistent inferences as to cause or mechanism. The term fell into disuse but was resurrected by the biologist E. O. Wilson in his 1998 book of the same name. Wilson uses the word to refer to the convergence of lines of argument from the diverse subdisciplines of the biomedical sciences *sensu lato*, including the cognitive sciences as well (particularly the cognitive processes of the scientist herself). In our bioinformatic context, the requirement is of the convergence of arguments at different *levels* of biological explanation, from the molecular right up through the ecological. It is not enough that patterns be found that relate the multiple levels, like Bowman’s fine examples (recently presented at these meetings) of covariance patterns relating SNP’s, clinical symptoms, and drug effects. It is required also that we be able to account for these patterns quantitatively, by properties of the biological systems that these drugs (for instance) are designed to alter.

More generally, to the consilience-attuned mind, the values of biological coefficients, such as physiological rate constants, need to be *parametrically* explicable by data deriving from instrumentation at other levels. The models of systems biology are composed of differential equations (partial differential equations, if we’re lucky). Claims that an intervention (such as a drug) can be expected to affect lumped observations upon such a system must involve the coefficients in those equations and the differential dynamics of their stable cycles. This is the domain of perturbation analysis and bifurcation analysis, not pattern analysis; it requires the measured values of biophysical constants, not estimated statistical quantities. *Models of the effects of interventions on organisms must be system-theoretic, not statistical.* Yet, perversely, we often systematically *remove* parametric information from our measured data, reducing microarrays to binary values, for example. The reason may be that we just cannot model the data as measured; but that is a simple confession of failure of the bioinformatic enterprise as biology. Inasmuch as the true reduced models are actually statistical-mechanical, based on continuous physical variables, no data transformation could possibly be more in opposition to appropriate principles of grounded numerical inference in this domain (systems biology) than that dichotomization. It is meretricious to promise applications in “drug design” in the absence of any method for optimizing the information content of a drug for an ongoing (living) organismal design.

Thus those six dogmas do not merely describe one critical approach to the rhetoric of statistical science per se; they also describe the ways in which that science can function to critique

the disciplinary rhetoric of the neighbors who supply its data or its justifications. All of the dogmas are strikingly absent from the majority of our LASR presentations — the *S* of *LASR* is being systematically suppressed. We must not depend for the justification of our investments of time on the fond hopes of our colleagues — we must rely on actual rules of method as statistical scientists teach them to other statistical scientists. Our work as statistical scientists needs to have solid intellectual merit in its rhetoric considered entirely on its own.

I challenge you, old friends and new colleagues, to try and find the six themes in the work that you are bringing to these LASR meetings, and, if you cannot find them, to consider whether that work truly comes under the heading of statistical science at all. It is necessary to find hypotheses that fit the data sets, and noise models that match prior knowledge of the instruments that helped us gather those data; and it is obligatory to show that no other explanations fit the data as well as our preferred mechanisms or causal narratives. Computations not in keeping with these desiderata are just that — mere computations, without passion, without insight, without real implications for human problems or human freedom. Good statistical science ought to touch on the dogmas — and not just one per paper, but more than one, wherever feasible. If I had more of your time than just today’s presentation — enough time to go over 450 manuscript pages of further examples with you — I might be able to convince you that all the finest examples of numerical inference across the natural sciences align with all six of these principles. If most are missing most of the time from the work we report at LASR, as I believe they are, the loss of gravitas — of intellectual weight — is our loss as well.

Acknowledgements. I am grateful to LASR for permitting me to mount its podium for an argument as iconoclastic as this. I have learned much from earlier conversations on these themes with Harald Martens, Norwegian University of Technology, and with Joseph Felsenstein, University of Washington. I am particularly grateful to another LASR regular, Clive Bowman, for years of discussions about the proper role of philosophy of science and philosophy of information in everything we statisticians do, from pharmacogenomics right across to fetal alcohol disorders. Parts of the argument here were presented to the conference “Measuring Biology” at the Konrad Lorenz Institute for Evolution and Cognition Research, Altenberg, Austria, September 2008, and have been previously published in volume 4 (2009) of the journal *Biological Theory*. The book *Numbers and Reasons* I mentioned in the second paragraph is currently under review by Cambridge University Press; if perchance you are among the reviewers of its current state, please be kind.

Literature Cited

Bookstein, F. L. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1991.

Bookstein, F. L. *Numbers and Reasons: Numerical Inference in the Sciences*. Book in preparation.

Edwards, W., H. Lindman, and L. J. Savage. Bayesian statistical inference for psychological research. *Psychological Review* 70:193–242, 1963.

Intergovernmental Panel on Climate Change [IPCC]. *Climate Change 2007: the Physical Basis*. Cambridge University Press, 2007.

Kuhn, T. S. The function of measurement in modern physical science. *ISIS* 52:161–193, 1961.

Mardia, K. V., F. L. Bookstein, J. T. Kent, and C. R. Meyer. Intrinsic random fields and image deformations. *Journal of Mathematical Imaging and Vision* 26:59–71, 2006.

Martens, H., and Tormod Næs. *Multivariate Calibration*. Chichester: John Wiley, 1989.

Pearson, K. Walter Frank Raphael Weldon, 1860–1906. *Biometrika* 5:1–52, 1906.

Peirce, C. S. *The Philosophy of Peirce: Selected Writings*, ed. J. Buchler. Routledge, 1940. The quotation is from page 151, where it is attributed to Peirce's Lectures on Pragmatism at Harvard in 1903.

Perrin, J. *Atoms*, 2nd English edition, revised. London: Constable, 1923.

Watson, J. *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. New York: Atheneum, 1968.

Watson, J., and F. H. C. Crick. A structure for deoxyribose nucleic acid. *Nature* 71:737–738, 1953.

Wigner, E. The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics* 13:1–14, 1960.

Wilson, E. O. *Consilience*. New York: Knopf, 1998.