areas can be highly misleading if viewed as quantitative evidence against a hypothesis?

I find, overall, that there is little I can add to the paper and little I can question. As I read the paper, I time and time again found myself saying—"That's a very good point; I couldn't agree more!"

ADDITIONAL REFERENCE

BERGER, J. (1987). Robust Bayesian analysis: Sensitivity to the prior. In *Foundations and Philosophy of Probability and Statistics, an International Symposium in Honor of I. J. Good on the Occasion of His 70th Birthday, May 25–26* (K. Hinkelmann, ed.). *J. Statist. Plann. Inference.* To appear.

Comment

David L. Banks

Jack Good's overview of the statistics/philosophy interface is delightful, informative and provocative. As usual, he combines substance with a great deal of engaging style and many scattered pearls. It is regrettable that his topic is so broad, for this sometimes forces him to treat major ideas with telegrammatic brevity; I hope that readers will be sufficiently intrigued to seek epexegesis in the references.

Over the years Good has started many hares at the border between statistical inference and the philosophy of science, and the article provides a partial synopsis of this facet of his research career. Although it is difficult for me to generate much disagreement with his principle views, I shall attempt to delineate aspects that make me either uneasy or eager for more development. Because the paper is rather a scattershot of topics, my comments are divided into thematic categories.

THE TYPE II WELTANSCHAUUNG

A major contribution is Good's development of dynamic probabilities. His overview emphasizes the relation between dynamic probabilities and partially ordered subjective probabilities, but I do not think his discussion carries the implications far enough. Good's point is that subjective probabilities change as one thinks, without new experimental information. In applications, one can only think so much, and thus one's subjective probabilities are necessarily approximate.

As an example, when someone states the Bieberbach conjecture, it sounds implausible and a good subjectivist might assign it a low probability. Further thought discovers numerous analytical functions that corroborate conjecture, inclining one to revise the probability upward. With a great deal of additional thought, a supremely clever person might rediscover de Branges' proof of the conjecture. Thus one's stated subjective probability depends on the amount of introspection spent upon the problem.

If a person is immortal, infinitely intelligent, perfectly sane (coherent) and reluctant to lose imaginary money, then she can construct an infinite sequence of hypothetical wagers that enables her to define her subjective probabilities with arbitrary precision. In practice, such perfect priors cannot be specified, and it behooves robust Bayesians to investigate the influence of errors induced by finite time, limited intelligence and insanity.

If error is caused only by Type II rationality (i.e., finite time), then it may be feasible to attempt a reasonably precise sensitivity analysis. For illustration, let's posit perfect intelligence and sanity, and assume that if one had infinite time, the prior chosen would be F. Let $\|\cdot\|$ be some reasonable metric on the space of measures (say $L^p[-\infty, \infty]$, $1 < p < \infty$), and take $\delta > 0$. Then one method of prior elicitation is to consider a sequence of distribution functions G_1, $G_2, \ldots$ such that for any $\delta > 0$ and any cdf H, there exists some n such that $\|G_n - H\| < \delta$ (on the line, one such sequence consists of step functions that place rational mass on the rational numbers; these are then ordered in analogy with Cantor's proof of the countability of the rationals). First one decides whether G_1 or G_2 is closer to one's prior with respect to the metric; then one considers each element of the sequence in turn, deciding whether the new element is closer to one's prior than the best cdf previously considered. After a fixed amount of time, one stops; let G_k denote the best cdf discovered, and G_m the last considered. Then F must lie in the region consisting of all cdfs closer to G_k than to $G_1, \ldots, G_m$. If one can search this region (and computer-intensive techniques are beginning to make this practical), then in principle one can either

- discover the prior that yields the most pessimistic analysis, or
- sample priors from the region and examine the distribution of inferences made from these.

David L. Banks is Assistant Professor of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.

Of course, this program may be difficult to implement, but it does suggest that one can incorporate the effect of finite time into the analysis.

A second sort of error is introduced by limited intelligence; it is related to Type II rationality in that if one had infinite time, then the mistakes could be discovered and corrected. These mistakes can take many forms (e.g., gross errors of the kind termed "blunders" by surveyors, as distinct from ordinary imprecision in measurement), but the most interesting relates to Good's assertion that the human mind cannot appreciate weights of evidence much smaller than a deciban. This graininess in the perceived weight of evidence imposes equivalence classes on the space of priors, so one cannot hope to specify one's beliefs too finely. One possible resolution is to use ε-contaminated priors of the sort discussed in Huber (1973) and Berger and Berliner (1986). Thus one makes the most intelligent decision one can about one's prior, and then incorporates additional uncertainty by permitting that prior to be contaminated with another cdf from some sensible set (perhaps the set of all measures, or all unimodal measures, etc.). The usual formulation requires one to specify the amount of contamination as ε, but it seems reasonable to borrow another tool from Good's tool kit and put a hyperprior over the parameter ε.

A third sort of error arises from the possibility of insanity. We all know highly intelligent people who are not strictly sane, and some are aberrant in rather subtle ways; thus a good Bayesian might put positive probability upon the proposition "I am not entirely sane" (this dovetails with Good's point that solopsism is not falsifiable; because one can imagine situations that corroborate solopsism, then a thoroughgoing Bayesian should give it weight). But if one admits the possibility of insanity, then the prior finally chosen may be entirely at odds with one's other beliefs; hence the stated prior may be arbitrarily distant from that obtained by a saner self. This type of error suggests that one must describe one's prior by a nonatomic measure upon the space of all measures, and the final analysis might resemble one based upon Ferguson's (1973) Dirichlet process formulation for nonparametric Bayesianity.

Good's general solution to these problems is to regard probabilities as partially ordered, taking values as intervals with vague end points. However, there are methods of analysis that explicitly respond to the three sorts of error listed above, and perhaps these enable a more structured analysis in practical problems. Surely increased attention should be paid to the influence of various sorts of error on approximate priors, and one hopes for the development of methods robust to all three sources of error. Also, one suspects that a thoughtful examination of Type II rationality could generate something like an uncertainty principle for Bayesian statistics, and this would be enormously interesting. *Wenige wissen, wieviel man wissen muss, um zu wissen, wie wenig man weiss.*

Incidentally, Good makes at least two testable statements at the border of psychology, philosophy and statistics. The first, already mentioned, is that a deciban is about the smallest weight of evidence humans can distinguish. The other is that people are better at assessing final probabilities than prior probabilities. One hopes that these intuitions will be followed up by some experimentalist.

DOOGIAN DOGMA

The partially ordered probabilities lead Good to sketch basic aspects of the Bayes/non-Bayes compromise or Doogianism. Unfortunately, the present treatment is so terse as to border on amputation, and I urge readers to refer at least to Good (1983f, Chapter 4). Basic tenets of Doogian dogma are:

1. In many cases, frequentist procedures can be used to approximate a Bayesian inference.
2. If a statistical procedure doesn't violate one's prior belief as developed under Type II rationality, then it can be used in an analysis.
3. Analysis implies paralysis (cf. *Hamlet*). If one thinks hard enough, then a corollary of Type II rationality is that anything you do must violate your beliefs.

The unhappy dilemma of the third point is salvaged in practice by the fact that one rarely has time or impetus to think really hard. So mainline frequentist procedures are often not excluded by the problem, and experience suggests that in many cases, they work well enough to serve. This is particularly true if one follows Good's lead in using frequentist methods to generate approximately Bayesian inferences.

However, this Doogian compromise seems somewhat *ad hoc*. Both Bayesian and frequentist procedures develop in natural and largely self-consistent ways, given separate starting points (i.e., as Good asserts, the archetypal Bayesian thinks he has point-valued prior probabilities, and the orthodox frequentist behaves as though all his priors have the interval value [0, 1], except for those used in modeling). But the Bayes/non-Bayes compromise position says that any analysis that is not contradicted by one's cursory examination is permitted. This does not generate a natural body of tests and procedures, which is unappealing. Also, different Doogians have even less certitude than Bayesians that the same experiment will lead to similar conclusions.

For example, let us assume that Jack Good has a secret monozygotic twin (which might explain his prolific output). They have spent their lives together, they have read the same books, and are in fact *zwei Seelen und ein Gedanke*. Specifically, they share the same subjective probabilities. Give them both the same data and the same problem. The Doogian formula implies that they will come to different conclusions for two different reasons.

1. IJG may have more time to introspect upon the problem that IJG′, so that the twins settle upon different interval-valued probabilities.
2. Even if the twins decide upon the same intervals, there may be several sorts of analysis that do not contradict these beliefs; thus IJG and IJG′ could pursue different analyses and reach different conclusions.

The first problem is intrinsic to the Type II view; the second problem is what seems most awkward for the Bayes/non-Bayes compromise. One needs a basis for deciding when enough thought has given to the problem, and a protocol for choosing among equally justified analyses.

CAUSALITY AND CASUISTRY

Probabilistic causality is an elegant piece of argument, and the concept seems potentially important, but it is not yet clear that the subject will be relevant to statisticians. After rereading the expanded treatment of the subject given in Good (1983f, Chapters 21 and 22), I am left with the fear that this is a case of *molto fumo e poco arrosto*. The philosophical side of the house seems very pleased with it, but one should note the historical tendency for a subject to split off from philosophy as soon as it becomes respectable (e.g., cosmology, mathematics, decision theory, etc.).

Good's rationale for the statistical value of probabilistic causality rests on the fact that $Q(E:F)$ agrees with a measure of association in contingency tables, and that it can be used to interpret the expected influence of a change in the regressor variables in linear regression. This does establish a connection with statistics, but I look forward to future developments in this area that will be more compelling. One possible application that enjoys the advantage of topicality is a plot over time of the estimated value of $Q(E:F)$, the tendency of smoking to cause lung cancer, based on the data available to the tobacco industry in 1945, 1950, . . . , 1985.

A similar comment applies to explicativity, in that it isn't clearly crucial to modern statistics. There is the possibility of important connections with model selection, and Good mentions the work of Akaike (1974) and Schwarz (1978), but the key comparisons have yet to be made. This is another area in which one hopes that Good's article will strike sparks.

ADDITIONAL REFERENCES

BERGER, J. and BERLINER, L. M. (1986). Robust Bayes and empirical Bayes analysis with ε-contaminated priors. *Ann. Statist.* **14** 461–486.

FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1** 209–230.

HUBER, P. (1973). The use of choquet capacities in statistics. *Bull. Inst. Internat. Statist.* **45** 181–191.

Rejoinder

I. J. Good

I'm most grateful to the discussants for their comments, both the generous and the critical ones. All four discussants seem to approve of some form of Bayes/non-Bayes compromise and with some other things I've said, but they have raised various issues that demand some response. I shall respond in the order in which the contributions are printed, but I deal first with probabilistic causality because three of the discussants have commented on it.

1. PROBABILISTIC CAUSALITY

The notation $Q(E:F)$ is an abbreviated notation, and, as I mentioned, the full notation mentions "the true laws of nature" (and other things) as *given* (to the right of a vertical stroke). When there are two or more scientific theories there will therefore be two or more estimates of $Q(E:F)$, and at most one of those estimates can be correct. This is my reply to one of the comments by Suppes where he described two different theories of learning only one of which can be (approximately) true.

I agree with Suppes's analogy with regression theory, in fact it is somewhat more than an analogy. If much of some other science is taken into account in a statistical or philosophical project, then the project is no longer regarded as just statistics or just philosophy. A physicist usually wants a better explanation of his data sets than can be provided by regression theory