squared modulus of the psi function. The theory amounts to a doctrine that there exist "systems" whose "state" can be described by a psi function satisfying certain rules of combination and of evolution in time. These "systems" relate to objectively describable repeatable experimental set-ups; and the theory is related to such set-ups mainly by interpreting the squared modulus of the psi function as a "long run frequency probability" over repetitions of such set-ups. No subjective element enters into this, although in relation to a single such set-up an observer may associate the quantum-theoretical probability with a subjective probability of the same magnitude. There are many fascinating puzzles here, well described by David Mermin in the April 1985 issue of *Physics Today*.

Like Good I see the future of the foundations of statistical inference in Bayes/non-Bayes compromises involving hierarchical models, objective data summarizations and in other directions. It is a pleasure to have been invited to discuss.

ADDITIONAL REFERENCES

BARNARD, G. A. (1987). R. A. Fisher—a true Bayesian? *Internat. Statist. Rev.* **55** 183–189.

PITMAN, E. J. G. (1965). Some remarks on statistical inference. In *Bernoulli, Bayes, Laplace* (J. Neyman and L. Le Cam, eds.) 215–216. Springer, New York.

Comment

James O. Berger

I recall being surprised upon first encountering the considerable interest of many philosophers in probability and statistics, interest at an often detailed technical level. Perhaps even more unusual is a serious professional interest in philosophy from a statistician or probabilist. Jack Good has had such a professional interest, virtually from the beginning of his career, and it is indeed a pleasure to view the world of "probabilistic philosophy" through his eyes.

One of the cornerstones of probabilistic philosophy was the development of the Bayesian and expected utility paradigms for processing information and making decisions. The paradigms were, however, an incomplete representation of reality, until Good incorporated the concept of partially ordered probabilities into their structures. I have written, in some depth, about this aspect of Good's work in Berger (1987), and so will refrain from further comments here.

I found Good's comment, that "... the future of statistics ... will be a compromise between hierarchical Bayesian methods and methods that seem superficially to be non-Bayesian," quite interesting. It is true that hierarchical Bayesian methods (including their empirical Bayes approximations) often have no workable classical analogues, and hence will be indispensable to the future of statistics; was more than this intended by the comment?

James O. Berger is The Richard M. Brumfield Distinguished Professor of Statistics, Department of Statistics, Purdue University, West Lafayette, Indiana 47907.

Isn't the left hand side of (2) often called a "weighted likelihood ratio"? I have several times been cynically amused that some statisticians will have no qualms about basing a decision on a weighted likelihood ratio with rather arbitrarily chosen weight functions, but will cry out in horror at the thought of using a Bayes factor with a prior that is actually thought about!

Another way of trying to understand the type of correction to a p-value given in (4), is to observe that, as long as N is at least moderately large,

$$\frac{p\text{-value}}{\text{Bayes factor}} \cong \frac{2\sigma g(\theta_0)}{\sqrt{N}[z + (.75)z^{-1}]};$$

here σ is the standard deviation of an observation, $g(\theta_0)$ is the value of the prior density as it approaches the null model θ_0 and z is the standardized (normal) test statistic $z = \sqrt{N}(\bar{x} - \theta_0)/\sigma$. Thus a p-value will behave roughly like a Bayes factor if it is multiplied by $\sqrt{N}$. (The above formula further suggests that multiplying p by $[z + (.75)z^{-1}]$ might be a beneficial standardization, but this is a comparatively minor additional correction.)

The idea of choosing a (perhaps crude) Bayes factor to be the significance test criterion certainly should be beneficial to classical testing. What, however, is the value of this to a Bayesian, who feels that all tail 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.