DeGroot, M. H. and Fienberg, S. E. (1986). Comparing probability forecasters: Basic binary concepts and multivariate extensions. In *Bayesian Inference and Decision Techniques* (P. K. Goel and A. Zellner, eds.) 247–264. North-Holland, Amsterdam.

Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates (with discussion). *Ann. Statist.* **14** 1–67.

Dickey, J. M., Dawid, A. P. and Kadane, J. B. (1986). Subjective probability assessment methods for multivariate-*t* and matrix-*t* models. In *Bayesian Inference and Decision Techniques* (P. K. Goel and A. Zellner, eds.) 177–195. North-Holland, Amsterdam.

Durbin, J. (1987). Statistics and statistical science. *J. Roy. Statist. Soc. Ser. A* **150** 177–191.

Evett, I. W. (1984). A quantitative theory for interpreting transfer evidence in criminal cases. *Appl. Statist.* **33** 25–32.

Finkelstein, M. O. (1978). *Quantitative Methods in Law.* Free Press, New York.

Fishburn, P. C. (1986). The axioms of subjective probability (with discussion). *Statist. Sci.* **1** 335–358.

Geisser, S. (1985). On the prediction of observables: A selective update (with discussion). In *Bayesian Statistics 2.* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 203–299. North-Holland, Amsterdam.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31** 423–447.

Huzurbazar, V. S. (1955). On the certainty of an inductive inference. *Proc. Cambridge Philos. Soc.* **51** 761–762.

Jeffreys, H. (1939). *Theory of Probability.* Clarendon Press, Oxford.

Jewell, W. S. (1974). Credible means are exact Bayesian for simple exponential families. *Astin Bull.* **8** 77–90.

Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S. and Peters, S. C. (1980). Interactive elicitation of opinion for a normal linear model. *J. Amer. Statist. Assoc.* **75** 845–854.

Kadane, J. B. and Winkler, R. L. (1988). Separating probability elicitation from utilities. *J. Amer. Statist. Assoc.* **83** 357–363.

Kahneman, D., Slovic, P. and Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases.* Cambridge Univ. Press, Cambridge.

Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin. English edition, Chelsea, New York, 1950.

Kuhn, T. S. (1974). *Structure of Scientific Revolutions.* Univ. Chicago Press, Chicago.

Lavis, D. A. and Milligan, P. J. (1985). The work of E. T. Jaynes on probability, statistics and statistical physics. *Brit. J. Philos. Sci.* **36** 193–210.

Lindley, D. V. (1972). *Bayesian Statistics: A Review.* SIAM, Philadelphia.

Lindley, D. V. (1984). Prospects for the future: The next 50 years (with discussion). *J. Roy. Statist. Soc. Ser. A* **147** 359–367.

Lord, F. M. and Novick, M. R. (1968). *Statistical Theories of Mental Test Scores.* Addison-Wesley, Reading, Mass.

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* Chapman and Hall, London.

Meinhold, R. J. and Singpurwalla, N. D. (1983). Understanding the Kalman filter. *Amer. Statist.* **37** 123–127.

Mosteller, F. (1988). Broadening the scope of statistics and statistical education. *Amer. Statist.* **42** 93–99.

O'Hagan, A. (1988). *Probability: Methods and Measurement.* Chapman and Hall, London.

Pratt, J. W., Raiffa, H. and Schlaifer, R. (1964). The foundations of decision under uncertainty: An elementary exposition. *J. Amer. Statist. Assoc.* **59** 353–375.

Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory.* Harvard Univ. Press, Cambridge, Mass.

Ramsey, F. P. (1931). Truth and probability. In *The Foundations of Mathematics and Other Logical Essays.* Kegan, Paul, Trench, Trubner, London.

Robinson, G. K. (1987). That BLUP is a good thing—the estimation of random effects. CSIRO, Melbourne.

Savage, L. J. (1954). *The Foundations of Statistics.* Wiley, New York.

Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton Univ. Press, Princeton, N.J.

Shafer, G. (1986). Savage revisited (with discussion). *Statist. Sci.* **1** 463–501.

Smith, C. A. B. (1961). Consistency in statistical inference and decision (with discussion). *J. Roy. Statist. Soc. Ser. B* **23** 1–37.

Sturrock, P. A. (1973). Evaluation of astrophysical hypotheses. *Astrophys. J.* **182** 569–580.

Wald, A. (1947). *Sequential Analysis.* Wiley, New York.

Wald, A. (1950). *Statistical Decision Functions.* Wiley, New York.

Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19** 326–339.

West, M., Harrison, P. J. and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *J. Amer. Statist. Assoc.* **80** 73–97.

Winkler, R. L. (1986). On "good probability appraisers." In *Bayesian Inference and Decision Techniques* (P. K. Goel and A. Zellner, eds.) 265–278. North-Holland, Amsterdam.

Zadeh, L. A. (1983). The role of fuzzy logic in the measurement of uncertainty in expert systems. *Fuzzy Sets and Systems* **11** 199–227.

Comment

George A. Barnard

The invitation to comment on Lindley's lectures arrived with a close deadline at a busy time. But, as always, his style is so clear and his thought so bold that I find the temptation to discuss at least some of his points irresistible.

When the University of Oxford—"the home of lost causes"—at last decided to set up a lectureship in mathematical statistics they called the resulting group LIDASE: the lectureship in the design and analysis of

George A. Barnard is Professor Emeritus at the University of Essex. His mailing address is Mill House, 54 Hurst Green, Brightlingsea, Colchester, Essex CO7 0EH, England.

scientific experiment. LIDASE eventually became the Unit of Biometry (with a Readership), then the Department of Biomathematics (with a Chair), and now the new Department of Statistics has Chairs in Applied Statistics and Statistical Science. While showing its traditional resistance to what might have turned out a passing fashion, the University lived up to its strong sense of history in choosing the first title. For the design and analysis of scientific experiment has formed the core study of mathematical statistics in this century and has been mainly responsible for its impressive growth. Building on foundations laid by others, our founding fathers—Gosset, Fisher, Neyman and Pearson (GFNP)—developed a theory of the design and analysis of experiments involving variable material. The wide applications these methods continue to find in agriculture, industry, medicine, biometry and other natural sciences laid the basis for a post-war "boom." More recently ideas evolved from these experimental methods have been applied in econometrics and in other social sciences.

Meanwhile two other traditions involving uncertainty continued older traditions of probability theory, especially in its applications to actuarial science and in the economic theory of risk taking. F. P. Ramsey and B. de Finetti saw themselves primarily as working within these traditions, although both saw clearly that their work had wider philosophical implications. Ramsey's brief references to Fisher's "likelihood" concept seem to me to make it clear that he did not see his work on subjective probability as having much immediate bearing on Fisher's ideas. There is certainly no evidence that he rejected the concept.

In this article Lindley presents a unitary concept of probability which attempts to encompass all forms of uncertainty wherever they arise. One may applaud the boldness of his attempt, while recognising, as does he, that there is a very long way to go before such a general theory can be regarded as complete. His Section 6 on the measurement of probabilities opens up very exciting possibilities. And his underlining of Geisser's arguments—and those of de Finetti—against the overuse of parametric formulations is welcome; though just as the impression sometimes given by GFNP that the introduction of prior distributional assumptions came close to mortal sin, the introduction of parameters when this is not strictly necessary does not always entail serious harm.

But while Lindley accepts, in Section 6, that we have a long way to go before we can claim to have a much broader calculus of uncertainty, some of his earlier criticisms of GFNP methods seem to assume we are farther ahead than in fact we are. For example, in relation to GLIM, Lindley criticises its emphasis on "likelihood ideas and ad hockeries." A Bayesian view would, he says, remove these. Until such a Bayesian view is forthcoming, free from ad hoc priors, one has leave to doubt. And more generally, one is entitled to ask whether his strictures on GFNP (no eighteenth century figure could represent them—perhaps Ben Franklin with his enthusiasm for experiment comes nearest) are justified. Within their narrower field, GFNP theories have a clarity and precision which suggests to me that when we approach the grand synthesis it will be found to incorporate the main principles advanced by GFNP.

The DASE ending of LIDASE serves to indicate the narrower view of the function of the GFNP statistician. Whereas Lindley is concerned to teach everyone how to face uncertainty, GFNP is concerned to help those prepared to undertake experiments with variable outcomes how to set up appropriate models, how to design experiments to test the appropriateness of these models and to learn to make them more specific, and then how to express in assimilable form the information supplied by the experiment. It will be for the client and anyone else who reads the results to supply their own background information insofar as this may be needed for a full interpretation.

A scientific experiment involves a repeatable set-up whose actual outcome x is one of a set $\{E\}$ of possibly observable outcomes. Associated with the experiment is a family $\{h(\theta)\}$ of relevant hypotheses, each of which specifies a probability function $p(\cdot \mid \theta)$ on $\{E\}$. It was Fisher's insight, arrived at over the period 1912–1924, to perceive that the primary inference, as he called it, from the observation x consists of the likelihood function $L_x(\cdot) = p(x \mid \cdot)$, defined on $\{h(\theta)\}$. Just as $p(\cdot \mid \theta)$, for each θ, defines a plausibility ranking on $\{E\}$, so $L(\cdot)$ defines a plausibility ranking on $\{h(\theta)\}$. They share the "product" property in relation to independent experiments. If x and y are possible results of two independent experiments, with probability functions $p(x \mid \theta)$, $q(y \mid \theta)$, the combined result $x.y$ (read: *x then y*) can be regarded as the result of a larger experiment, with associated probability function $p(x \mid \theta)q(y \mid \theta)$; the likelihood based on $x.y$ is $L_{x.y}(\theta) = L_x(\theta)L_y(\theta)$. The second experiment may sometimes be dependent on the result of the first. In such a case $q(y \mid \theta)$ must represent a conditional distribution, and should be written $q(y \mid \theta, x)$.

The contrast between a result x of a specified experimental set-up and an observed fact F such as the existence of a cluster of childhood leukemia cases in the neighbourhood of a nuclear installation can be seen from the discussion of the latter reported in (Gardner et al., 1989). F arose as the result of a visit to the installation by an investigator looking for cancer cases among the adult employees of the installation. If a knowledgeable person had attempted to list the possible results of the investigator's visit, it is most unlikely that F would have been listed among them. (Oddly enough, a less than knowledgeable person just possibly might have listed F; cf. Gardner et al., 1989). There would thus be no clearly defined set corresponding to $\{E\}$ in this case. I am not by any means saying that statisticians have no role in discussing issues raised by facts such as F; indeed, as the JRSS discussion shows, they have much to contribute—not least a habit of dispassionate adherence to verifiable fact which their GFNP experience encourages.

Because, with a scientific experiment, it is possible to list the possible outcomes, and to change the set-up so as to cease to distinguish between two given possible outcomes, the logical operations of disjunction, $\vee$, and negative, $\sim$, can meaningfully be applied to the possible results. $(x \vee y)$ denotes a result of an experiment in which the set-up fails to distinguish x from y. $(\sim x)$ denotes the contrary to x in an experiment in which all that is to be observed is whether or not x happens. The addition and the complementation laws of probability therefore apply to possible experimental results. It is otherwise with possible hypotheses $h(\theta)$. Merely to deny $h(\theta)$ does not usually tell us what the true distribution is. And the disjunction $h(\theta) \vee h(\theta')$ merely tells us that the probability of x is either $p(x \mid \theta)$ or $p(x \mid \theta')$; it does not tell us which of these it is. Thus the addition and complementation rules of probability do not directly apply to likelihood.

Although likelihood has no use for the addition rule, the numerical value of a log-likelihood ratio, $\Lambda(\theta, \theta') = \mid \ln L_x(\theta)/L_x(\theta') \mid$ has a direct interpretation as indicating the strength of the evidence for (say) θ versus θ'. Because if θ' were true, and the mean value of the log-likelihood ratio from a single observation was $\lambda(\theta, \theta')$, then the ratio Λ/λ would indicate how many more observations would be likely to be needed to reverse the plausibility ranking. And we may see here how this part of GFNP would fit into the global theory of uncertainty at which Lindley aims by noting that if the global theory enabled You to assess the log-odds in favour of θ' as against θ as Λ, the data x would just serve to render You neutral as between these two alternatives.

It may be worthwhile drawing attention to the fact that in a letter dated 3 April 1922 G indicates to F that he (G) regards likelihood as the posterior relative to a uniform prior; and there seems to be no evidence that he ever came to think otherwise. With his usual percipience he noted that if you are putting together two posteriors from independent experiments, each with its prior $\pi(\theta)$, $\pi'(\theta)$ it is only when you take $\pi(\theta) = \pi'(\theta) = C$ that you get consistent results.

It was observed by Fisher that it is sometimes possible for the set $\{E\}$ of possible results of a single experiment to be mapped 1–1 onto a product set $\{G\} \times \{K\}$, so that x is mapped to (u, v), with $u \in \{G\}$ and $v \in \{K\}$ while $p(x \mid \theta)$ factors into $q(u \mid \theta) r(v \mid \theta, u)$ with $q(. \mid \theta)$ the same for all θ. It may then be possible to regard the experiment as equivalent to two experiments, the first with possible result set $\{G\}$, the second with possible result set $\{K\}$; and the set-up of the second experiment may be partly determined by the result of the first. In such a case u is ancillary, and the relevant likelihood is that derived from the conditional distribution of v, given u.

An example considered by Laplace may serve as an instance: We have a set x of three measurements (x_1, x_2, x_3) of a quantity each measurement independently liable to an error e with density $\frac{1}{2} \exp - \mid e \mid$. If θ is the true value the joint density of the x's is

$$\text{(1)} \quad p(x \mid \theta) = (1/8)\exp - \{\mid x_1 - \theta \mid + \mid x_2 - \theta \mid + \mid x_3 - \theta \mid\}.$$

By a permutation P, one of six equally probable such, we can arrange the x's in ascending order, and then we can denote the middle x by v, while the lowest x is $v - u_1$, and the highest is $v + u_2$. The Jacobian of the transformation from x to (P, u, v) is $1/6$ and the joint density of (P, u, v) is therefore

$$\text{(2)} \quad q(P, u, v \mid \theta) = (1/48)\exp - \{\mid v - \theta - u_1 \mid + \mid v - \theta \mid + \mid v - \theta + u_2 \mid\}.$$

The probability function of P is $1/6$, constant over $\{\theta\}$. So also is the probability function of u—obtained by integrating out $(v - \theta)$ from q; we are therefore left with the observation v with density

$$\text{(3)} \quad r(v \mid \theta, u, P) = K(a, b)\exp - \{\mid v - \theta - a \mid + \mid v - \theta \mid + \mid v - \theta + b \mid\}$$

where the normalising constant K depends on the observed values $(a.b)$ of (u_1, u_2). The three observations are equivalent to a single observation v whose error $v - \theta$ has the density (3).

Perhaps a more instructive example is provided by the 2×2 table where we have m observations, each 0 or 1, from population I and n observations, each 0 or 1 from population II, with $\theta = (p_1, p_2)$ representing the probabilities of 1 in I and in II, respectively. A reduction similar to that illustrated above shows than we can represent x as (a, c) with a the number of 1's from I and c the number of 1's from II. The log-likelihood is $\ln L_x(\theta) = a(\ln p_1 - \ln q_1) + m \ln q_1 + c(\ln p_2 - \ln q_2) + n \ln q_2$. If, as often, we are interested only in the "difference" between p_1 and p_2, we have to determine how this "difference" is to be expressed. It turns out, in the common case where $m = n$, that putting $\xi = \frac{1}{2} \ln(p_1 q_2/p_2 q_1)$, the semi-log of the odds ratio, and $\eta = \frac{1}{2} \ln(p_1 p_2/q_1 q_2)$, the log-likelihood becomes

$$L_x(\xi, \eta) = (a+c)\eta + (a-c)\xi - m\ln\{1 + \exp 2\eta + \exp \eta(\exp \xi + \exp - \xi)\},$$

the sum of three terms: the first depending on η not on ξ, the second on ξ alone, and the third depending mainly on η but also on $|\xi|$, not on ξ itself. This suggests representing the data as (1) an experiment with $m = n$ fixed, in which $r = a + c$ is observed, followed by (2) an experiment in which $a - c$ is observed, the value of r from (1) being fixed. The first experiment gives us no information at all about the sign of ξ, and only a small amount of information about its magnitude, extractable only if some further information about the value of η can be introduced. In the absence of such information, and when (as often) the sign of ξ is regarded as perhaps more important than its numerical value, it is therefore reasonable to ignore (1) and to derive inferences about ξ from (2) alone. This leads to Fisher's conditional analysis of the 2×2 table.

Fisher's test for the 2×2 table is best regarded from the point of view suggested by NP in 1933. For them, "testing" the hypothesis H amounted to fixing a set C of values of x such that, if x fell in C, H would be rejected, not otherwise. If instead of noting the actual value of x, all that was observed was whether or not x fell in C, the resulting likelihood function—the "power function," as NP called it—would most nearly match the original power function if C consisted of just those values of x for which the likelihood ratio for an alternative to H exceeded some fixed value. In addition to its interpretation as a likelihood function, the power function had its well known interpretation in terms of long run risks of error—especially important in the industrial applications with which P was much concerned. The only weakness of the NP approach in this problem was, that it failed to note the dependence of the likelihood function on r.

Precise definition of the likelihood function requires precise specification of the probability function $p(x \mid \theta)$. It is often possible to do this exactly or to an excellent approximation when $\{E\}$ is discrete; but when $\{E\}$ is continuous we need a form of reasoning which does not require such precise specification. We can sometimes do this by regarding L as known only to within some small ε, but it is better to proceed otherwise by adopting a *pivotal model.* This is always possible in the continuous case. We can illustrate by generalising the Laplace example considered above. The basic pivotal here is $\mathbf{p} = (p_1, p_2, p_3) = (x_1, - \theta, x_2 - \theta, x_3 - \theta)$, whose density we denote by $\varphi(p_1, p_2, p_3)$. We can transform $\mathbf{p}$ to $(P, u_1, u_2\ v)$ by the same transformation as used before, with the same result in the case of P, u_1, and u_2, but now $v = \tilde{x} - \theta$ where $\tilde{x}$ is the median of the x's. If P, a and b are the observed values as before, we can imagine ourselves learning these before we learn the value of $\tilde{x}$ and again this justifies us in taking v to have its conditional density

$$\Psi(v) = K(a, b)\varphi(v \to a, v, v + b).$$

Our information about θ is again equivalent to that provided by a single observation, now $\tilde{x}$, subject to an error v having the density $\Psi(v)$. It is now a mathematical problem, given an approximate specification of φ, to derive an approximate specification of Ψ. It often turns out—depending on a and b—that we can specify Ψ more accurately than we can specify φ, in which case we say we have a robust sample.

Thus far I have been stressing the extent to which GFNP and pivotal inference preserve likelihood functions and, to this extent, must be compatible with Bayesian reasoning. I now come to a point where, it seems to me, the GFNP approach can improve on Lindley's programme. Abstaining from unnecessary generality, suppose we have a sample of size $n \geq 2$ with mean $\bar{x}$ and standard deviation s from a normal population with unknown mean μ and unknown standard deviation $\sigma = \exp \zeta$. The two linear pivotals

$$t = (\bar{x} - \mu)\sqrt{n}/s \quad \text{and} \quad z = \ln s - \zeta$$

have a well known joint density $\varphi(t, z)$. They are jointly sufficient for (μ, ζ) (or equivalently, for (μ, σ)) because the likelihood function $\varphi(t_0, z_0)$, obtained by subsituting the observed $\bar{x}$, s in t and z is equivalent to the likelihood function based on the original observations. (See Sprott, 1989, and references there given, for more on linear pivotals.) *If nothing else is known about* ζ, we may take the marginal density of t—Student's density on $n = -1$ degrees of freedom—as conveying all the available information about μ. From a Bayesian point of view, this is justified by saying that the "ignorance" priors for both μ and ζ are locally uniform, deducing the joint posterior density for (μ, σ), and then marginalising for the density of μ. From the pivotal point of view, however, we can simply say that our ignorance of ζ and of μ implies we can have no grounds for supposing, after the data are known, that the still unknown values of (t_0, z_0) are anything other than a randomly selected pair from the joint distribution $\varphi(t, z)$; consequently the appropriate distribution for t is its marginal distribution. (If the value of ζ were known, of course, the resulting distribution for t would instead be its conditional distribution, given the now known value of z; while if ζ were simply known to lie between limits ζ_1, ζ_2, the relevant density of t would be obtained by integrating φ with respect to z from $z = \ln s - \zeta_2$ to $z = \ln s - \zeta$; see Chamberlin and Sprott, 1989.) The "inference from ignorance" to the joint distribution of (t_0, z_0) seems to be more direct than that assuming "ignorance" priors; and it has the advantage of being compatible with a frequency interpretation for μ-interval statements derived from t.

As my final example, I take the Behrens–Fisher problem (which also is easily generalised to nonnormal densities). We have $(m, \bar{x}, s)$ from (approximately) $N(\mu, \sigma')$ and $(n, \bar{y}, s')$ from (approximately) $N(\mu + \delta, \sigma'')$ and we wish to estimate δ. (Perhaps I should interpolate here that I agree with Lindley's point about estimation *not* consisting of "point" and "interval" estimation. Sprott (1989) has suggested the useful term "inferential estimation" to denote what we are now discussing.) That is, we wish to find a pivotal $G(\bar{x}, \bar{y}, s, s', \delta)$ whose distribution is (approximately) known and which could either serve to test hypotheses about the value of δ or could be combined with prior information to derive a posterior distribution.

To begin with it is convenient to transform the scale parameters to (σ, γ) where

$$\sigma^2 = \frac{\sigma'^2}{m} + \frac{\sigma''^2}{n} \quad \text{and} \quad \gamma\sigma^2 = \left(\frac{\sigma'^2}{m}\right)$$

so that $(1 - \gamma)\sigma^2 = (\sigma''^2/n).\sigma^2$ is the variance of the difference $\bar{y} - \bar{x}$, and $\gamma : 1 - \gamma$ is the ratio in which this variance is shared between $\bar{x}$ and $\bar{y}$. Then the basic pivotals are reducible to four:

$$t_x = (\bar{x} - \mu)\sqrt{m}/s, \quad t_y = (\bar{y} - \mu - \delta)\sqrt{n}/s',$$
$$z_1 = s/\sigma\sqrt{m\gamma} \quad \text{and} \quad z_2 = s'/\sigma\sqrt{n(1-\gamma)}.$$

It is shown in Barnard (1982) that any function $H(t_x, t_y, z_1, z_2)$ which is of the form $G(\bar{x}, \bar{y}, s.s', \delta)$ can take at most three distinct values. So our problem is incapable of solution as stated. However taking γ as known allows it to be a fifth pivotal, and we can find a nontrivial $H(t_x, t_y, z_1, z_2, \gamma)$ of the required form G. In fact, the quantity

$$t = \frac{\bar{y} - \bar{x} - \delta}{\sqrt{\dfrac{[(n-1)s'^2/(1-\gamma)] + [(m-1)s^2/\gamma]}{m+n-2}}}$$

has Student's distribution on $m + n - 2$ degrees of freedom and as a linear pivotal the associated likelihood is equivalent to the marginal likelihood of δ. Further, we have a pivotal for γ,

$$f = z_2/z_1 = s'\sqrt{m\gamma}/s\sqrt{n(1-\gamma)}$$

distributed as a multiple of $\sqrt{F}$ on $(n - 1, m - 1)$ degrees of freedom. We could use this to determine plausible limits for γ, and then derive the associated range of t values corresponding to any proposed value of δ. But a preferable way would be to assess a prior distribution for γ, use the likelihood based on f to deduce a posterior distribution for γ, and thence deduce the density of t as averaged over this posterior distribution. The likelihood based on f has a Beta form, so that it is easy to programme the posterior of γ—and the derived distribution of t—relative to a γ prior of the Beta form $K\gamma^a(1 - \gamma)^b$.

A widely accepted Bayesian solution of this problem, due originally to Jeffreys, assumes for the scale parameters of σ', σ'', a joint prior density element $d\sigma' d\sigma''/\sigma'\sigma''$, indicating ignorance of both scale parameters. This corresponds to taking for γ a Beta prior with $a = b = -1$, with divergence at both $\gamma = 0$ and at $\gamma = 1$. But this is highly implausible. We shall almost certainly have arranged the sample sizes m, n so as to make the variances of $\bar{x}$ and $\bar{y}$ at least of the same order of magnitude, and the observation of a value of $s'\sqrt{m\gamma}/s\sqrt{n(1-\gamma)}$ reasonably distant from 0 and from 1 would contradict the assumption $a = b = -1$. Jeffreys appears to have overlooked the fact that the independence implied by the product form of the probability element implies not only ignorance of both σ' and σ'', but also that even if σ' were known exactly we would still be entirely ignorant of σ''. In another connection, Jeffreys proposes for a parameter like γ a prior with $a = b = -\frac{1}{2}$, which still has infinite density at each end of the range, but which does at least converge. In view of the possible adjustment of m and n already mentioned, I would suggest that $a = b = 0$—a uniform prior between 0 and 1—would often be closer to the truth; while when the problem arises in another context (that of recovery of interblock information) it is likely that past experience will be available to suggest perhaps unequal values of a and b. The ease of programming with a and b as adjustable parameters means that when in doubt we can try a range of (a, b) values; and if—as will often happen—varying these over a plausible range makes little difference to the distribution of t, we need not bother to examine the matter very closely.

I have sketched a range of GFNP approaches to various problems to bring out two main points: (1) the methods aim for the most part to reduce the original data to a readily assimilable form in such a way as to preserve the likelihood function. Any individual reading the reduced forms can therefore add his or her own prior distributions, if desired. There is no basic incompatibility with a Bayesian approach. (2) Those of us who prefer, whenever possible, to think in terms of likelihood, or of pivotal inference with its "coverage frequency" interpretations are free to do so. One of Lindley's most valuable contributions has been his propounding of the "Cromwell principle," which I like to think of as expressing the fact that we should always bear in mind that we may have overlooked some important possibility. Insofar as adoption of a Bayesian prior requires us to distribute the total unit probability over a specified set of possibilities, it has always seemed to me that such a procedure is incompatible with the Cromwell principle. The likelihood function serves to rank the possibilities we have contemplated in order of plausibility on the data, while leaving entirely open the possibility that other possibilities, not thus far contemplated, exist. And the "long run frequency" interpretations of pivotal inferences offer a degree of objectivity about our inferences which is also a valuable feature. Such long run interpretations are, of course, conditional on the acceptance of the model assumptions—including, as in the Behrens–Fisher example, any assumptions we choose to make about the long run frequency of occurrence of parameter values such as those of γ.

Such model assumptions need to be checkable. And here comes the first strictly non-Bayesian point: in discussion of Box (1980), Lindley introduces the idea of a "small world" within which You can construct a model for Your beliefs. The "small world" is then regarded as part of a "larger world" within which the small world is to be tested. "We may be surprised if $p(x \mid \theta)$ is tiny within the small world but unless there is a θ' with $p(x \mid \theta')$ bigger within the larger world, then θ must remain a plausible value and the surprise must be accepted." The assumption here is that "there is not" and "we cannot imagine" a θ' *for which we can calculate* $p(x \mid \theta')$ are equivalent. But when, in 1887, Michelson and Morley performed their experiment, their null result was very surprising—so much so that it was often repeated, with many refinements. For 10 years no one could imagine a θ'. Then Lorenz and Fitzgerald suggested their contraction as a possible θ'; but repetitions with varying framework materials rendered this implausible, and it was not until 1905 that Einstein's theory provided a θ' which slowly won wide acceptance. Again, it was not until the general theory of 1915 that a tenable θ' was provided for the motion of the perihelion of Mercury which had been noted more than 50 years earlier by Leverrier and for which attempts to find a "Vulcan" had long since failed. In both these cases, I would suggest that over a period of years the relevant θ was *not* regarded as plausible, and yet there was no θ' with a *calculable* $p(x \mid \theta')$. As of this writing the discussion on leukemia clusters referred to above provides another example, though the suggested association with movements of population may soon provide some approximation to such a $p(x \mid \theta')$.

The repeatability of scientific experiments is important here. Two independent sets of physicists find surprising evidence of cold fusion. Immediately others attempt to repeat. Had the attempts succeeded "acceptance" of surprise would not have been an option. There would in that case have soon—though not immediately—been suggested θ''s with $p(x \mid \theta') > p(x \mid \theta)$. Such suggestions would have been tentative until verified by further experiments, which would soon have been forthcoming. The fact that the "cold fusion story" lived for months, while the Michelson–Morley story went on for decades serves to illustrate the vast increase in the amount of experimental research there has been. But these times of transition, short though they may now be, are the growing points of scientific knowledge and it is in them that statistical methods, including the much maligned P values, play a specially important role.

On a point of detail: the example of Basu which Lindley quotes as an indefensible ancillary does not fit with the account of ancillarity given above. One can transform the space of x into the set of integers, and the interval [0, 1); but neither of the associated sets of probability is independent of θ. On the other hand when the experiment consists in n repetitions x_i of an observation, we can always imagine the result obtained by a preliminary determination of n, followed by a permutation giving the observed order of the x_i, followed by the ordered sample. If the first two parts have distributions independent of θ they qualify as ancillaries. This accounts for the common choice of a fixed n sample space though, as the 2×2 case illustrates, further reductions may be desirable.

On a point of history: in a certain sense young Berkeley was a wayward son of GFNP, born around 1934, as a result, I conjecture, of a liaison with A. N. Kolmogoroff. It will be remembered that ANK published at that time a book, *Grundlagen der Wahrscheinlichkeitsrechnung*, which served to change what had been a dubious area of study into a perfectly respectable and richly structured branch of pure mathematics. The N component of GFNP was much impressed with ANK, having himself belonged to that extraordinarily brilliant group of pure mathematicians who grouped themselves around the journal *Fundamenta Mathematicae*. It was apparent to N that while probability theory had received definitive rehabilitation, the same could not be said of the other measure of uncertainty which N, as well as the other parts of GFNP, had been calling likelihood. N therefore determined to rethink GFNP in purely probabilistic terms while adhering to the doctrine which he had learned from P, that prior distributions were heresy. He soon realised that if one stuck to probability only, one would have to abandon any idea of statistical *inference* and speak only of *inductive behavior*. Here one defined in full rigour the mathematical model of one's experiment (it was at this time that N began to lay down that the set of possible results was required to be a measure space) and, in advance of observing the results, one laid down the probabilistic properties required of those functions of the data that one intended to calculate. Thus one might require a "point estimate" of θ to be "unbiased," or require an advance specified upper bound to the probability of error of the first kind. That such requirements could lead to absurdities such as randomised "conclusions," assertions with only 90% confidence that a real number lay between $-\infty$ and $+\infty$, etc., was noted by several of the older Berkeley's associates; but the energy, courage, generosity of spirit, brilliant wit, and human warmth of N's character so impressed all those who came into contact with him that the inherent impossibility of the task N had set himself was not stressed, and old Berkeley grew into the over-rigid system which Lindley so mercilessly attacks. Of course, as with the somewhat similar attempts by von Neumann, Birkhoff, and others to "pure-mathematicise" quantum physics, N's programme produced many insights and valuable results in spite of its ultimate failure.

ADDITIONAL REFERENCES

BARNARD, G. A. (1982). A new approach to the Behrens-Fisher problem. *Utilitas Math.* **21B** 261–271.

BARNARD, G. A. (1989). On alleged gains in power from lower *P* values. *Statistics in Medicine.* To appear.

CHAMBERLIN, S. R. and SPROTT, D. A. (1989). The estimation of a location parameter when the scale parameter is confined to a finite range: The notion of a generalised ancillary statistic. *Biometrika* **76** 609–612.

GARDNER, M. J. et al. (1989). Cancer near nuclear installations. *J. Roy. Statist. Soc. Ser. A* **152** 305–384.

SPROTT, D. A. (1989). Inferential estimation, likelihood, and linear pivotals. *Canad. J. Statist.* To appear.

Comment

James O. Berger

1. INTRODUCTION

There are many reasons to adopt the Bayesian paradigm. Professor Lindley emphasizes the foundational and axiomatic rationales in this paper. Having followed that route to Bayesianism myself, I am particularly appreciative of the job Lindley has done in illuminating the route. I only regret that this paper was not around when I started studying the issues.

I emphasize the foundational nature of Lindley's paper for two reasons. First, it is a common misconception that the arguments for Bayesian statistics are all theoretical, as opposed to practical. To the contrary, an extremely strong case for Bayesian statistics can be made purely on the pragmatic grounds that it is much easier to understand and yields sensible answers with less effort. Lindley has reasonably concentrated on the foundational side, but it is important to note the existence of these very pragmatic rationales. Of course, I completely agree with Lindley that foundational issues can have a profound effect on practice.

The second reason for mentioning the foundational nature of the paper is that, in foundational matters, virtually everyone disagrees in some respect, even (or perhaps especially) Bayesians. Thus the bulk of my discussion focuses on the foundational differences that I have with Lindley, primarily the issue of specification of unique prior probabilities. While this is perhaps a significant issue foundationally, it is much less of an issue in terms of Bayesian statistical practice. Hence, my disagreements with Lindley are actually quite minor from the perspective of statistics in general. Indeed, my motivation for raising the issue (in Section 3) is mainly to argue that uncertainty in probability specifications can be incorporated into the Bayesian paradigm without any major changes being necessary.

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

2. FREQUENTIST BAYESIANISM

As I read Section 1 of the paper, I agreed with virtually all of the points raised but felt uneasy at the conclusion that coherence is missing from the Waldian paradigm. After all, admissibility is at the heart of the paradigm and, in a sense, admissibility is just a frequentist version of coherence.

Would Wald have disagreed that the correct solution to the mixture problem is to choose a procedure that is Bayesian? Perhaps not. Indeed, there have existed frequentists who consider themselves coherent Bayesians, in the sense that they agree with the use of Bayes' rules, and even utilization of prior information, but still want to base their evaluations of accuracy on frequentist (Bayesian) measures of performance. Such statisticians would presumably disagree with Lindley's statement that "only the Bayesian attitude is coherent . . . Consequently the sample space is irrelevant." They would agree with the first part, but disagree with the second because of their insistence that only frequentist measures are meaningful.