New Perspectives on Keynes

Annual Supplement to Volume 27

History of Political Economy

Edited by Allin F. Cottrell and

Michael S. Lawlor

Duke University Press

Durham and London 1995

Runde, J. H. 1994. Keynes after Ramsey: In Defence of *A Treaty on Probability*. *Studies in History and Philosophy of Science* 25:97–121.
Schmeidler, D. 1989. Subjective Probability and Expected Utility without Additivity. *Econometrica* 57:571–87.
Shackle, G. L. S. 1972. *Epistemics and Economics: A Critique of Economic Doctrines*. Cambridge: Cambridge University Press.

Jeffreys, Fisher, and Keynes: Predicting the Third Observation, Given the First Two

Teddy Seidenfeld

During three years, from 1932 through 1934, the *Proceedings of the Royal Society of London* was the setting for a stimulating, five-article exchange between Harold Jeffreys and Ronald Fisher about their differing views on the foundations of statistical inference. In what surely was a rare event in any debate with Fisher, Jeffreys got the first and the last word (Jeffreys 1932, 1933, 1934; Fisher 1933, 1934).[1] For our purposes at this conference on Keynes, I propose that we examine how, starting with Jeffreys's first rebuttal (1933) to Fisher, and continuing through Fisher's second reply (1934), and on to Jeffreys's final rebuttal (1934), each side used Keynes's 1921 *Treatise on Probability* to argue that the other was committing a foundational error.[2] To do that, first I review the statistical arguments Jeffreys and Fisher set out in their initial papers in this sequence. Then I examine how each side tried to co-opt Keynes's theory. Last I indicate some contemporary work that reflects, to my mind, how one aspect of the debate has evolved over sixty years.

I thank Rob Kass for some helpful comments on the material presented here.

1. However, just one year later, Fisher did not pass up the advantage of the "last reply" in his 1935 presentation and discussion, "The Logic of Inductive Inference," in the *Journal of the Royal Statistical Society*.

2. For a different perspective on this Jeffrey-Fisher exchange, see David Lane's stimulating essay from 1980.

Inverse Inference According to Jeffreys and to Fisher

In the 1930s, and even today, the foundational litmus test of a statistical theory is its solution to inverse inference: inference from "sample" to "population." A textbook case will serve as our heuristic. Let $(x_1, x_2, \ldots, x_n)$ be n i.i.d. $\mathbf{N}(\mu, \sigma^2)$ observations, where both parameters are unknown. What does your favorite statistical theory authorize may be inferred about the normal mean μ and variance σ^2 of the "population" from which the n observations have been independently sampled?

Harold Jeffreys was an advocate of Bayesian statistical inference, which solves inverse inference according to Bayes's rule:

$$P(\text{Hypothesis} \mid \text{Data}) \propto P(\text{Data} \mid \text{Hypothesis}) \times P(\text{Hypothesis}),$$

or,

$$\text{Posterior probability} \propto \text{Likelihood} \times \text{Prior probability.}$$

For our heuristic example, this becomes (in densities):

$$p(\mu, \sigma \mid x_1, x_2, \ldots, x_n) \propto (\sigma^{-n}) \exp[-\{(n-1)s^2 + n(\bar{x} - \mu)^2\}/2\sigma^2] p(\mu, \sigma)\, d\mu\, d\sigma,$$

where the sample variance $s^2 = \sum_i (x_i - \bar{x})^2/(n-1)$ and the sample average $\bar{x} = \sum_i x_i/n$, are jointly sufficient. But what is the joint prior probability for the unknown parameters, μ and σ?

Jeffreys (1931), in *Scientific Inference*, had already argued that, in estimation, the appropriate probabilistic representation of ignorance about a parameter depends on how that parameter functions in a statistical model of the data to be acquired. That is, rather than using a Laplacean, uniform prior to depict prior ignorance about a parameter, Jeffreys argued by way of statistical symmetries what that ignorance prior should be. For inference about a location parameter—for example, for inference about the normal mean μ (given the variance, σ^2)—he argued that shift-invariance for the sample average justified the uniform (improper) prior density $d\mu$.[3] Likewise, for inference about a scale parameter—for example, for inference about the normal variance σ^2 (given the mean, μ)—he argued that invariance for powers justified the (improper) prior density, $d\sigma/\sigma$.[4]

3. Note that $\bar{x}$ is sufficient for μ, given σ^2.
4. Note that $\sum_i (x_i - \mu)^2$ is sufficient for σ^2, given μ.

However, Jeffreys found no compelling reason to mandate a joint ignorance prior for the two parameters (μ and σ) that is merely the product of the separate ignorance priors, $p(\mu, \sigma) \propto 1/\sigma$.[5]

The first of the five papers, Jeffreys's 1932 essay, begins with a novel argument for the improper joint density, $d\mu d\sigma/\sigma$. His reasoning is elegant. Let (x_1, x_2, x_3) be three (continuous) i.i.d. random quantities from a statistical model. Let H be the hypothesis that x_3 lies (strictly) between x_1 and x_2. *Prior to observing the data*, the probability of H is 1/3. (It is immediate from the assumption the data are i.i.d. that all six orders are equiprobable; hence, for two of six equiprobable cases H obtains.) Suppose we observe two values (x_1, x_2) of three i.i.d. $\mathbf{N}(\mu, \sigma^2)$ variables. Jeffreys asked: What prior probability over the unknown parameters (μ, σ) leads to the conclusion that $P(H \mid x_1, x_2) = 1/3$, regardless of the specific values of (x_1, x_2)?

Expressed somewhat differently, Jeffreys's question is: What prior probability on the parameters preserves the ignorance we have initially about the relative order statistic for x_3 with respect to x_1 and x_2, regardless of the observed values of x_1 and x_2? The answer is, of course, the joint (improper) prior proportional to $1/\sigma$.[6] So, Jeffreys had a new reason for the two-parameter "ignorance" prior he used in Bayes's rule, based on a constraint for predictions.

R. A. Fisher was no Bayesian—not in the 1920s when he helped to create the foundations of significance testing, maximum likelihood, and the theory of statistical estimation and not in the 1930s when he set the foundations for randomized experimental design; nor was Fisher Bayesian in his many presentations (beginning in 1930 and ending only with his death in 1962) of his "fiducial" solutions to inverse inference. By 1933, Fisher's *Statistical Methods for Research Workers* was in its fourth edition and Bayes's rule was *not* one of the tools in the toolbox that is *SMfRW*! In 1933, Fisher was enjoying his newest invention, the enigma of fiducial probability. Neyman-Pearson hypothesis testing was a half-

5. The problem followed Jeffreys through much of his career. See, e.g., section 3.10 (especially p. 182) in *Theory of Probability* ([1961] 1967) to see the tension between Jeffreys's Invariance Theory applied to prior ignorance for μ and σ jointly versus separately.
6. In section 3.8 of *Theory of Probability* ([1961] 1967), Jeffreys shows that for a location-precision model, with parameters (α, h), respectively, the joint prior expressed as $d\alpha dh/h$ yields the desired probability $P(H \mid x_1, x_2) = 1/3$. (Precision is the multiplicative inverse of the scale parameter.) However, Jeffreys is unable to show the converse—that is, there remain some open cases whether this prior is unique for all location-scale families.

dozen years old, and confidence interval theory (according to Neyman, that is) was still a year away.[7]

Thus, I speculate it came as no surprise to Jeffreys that Fisher was unwilling to accept the novel derivation of Jeffreys's joint "ignorance" prior for the two normal parameters. For Fisher, as for many non-Bayesians, each probability assertion—whether as prior, likelihood, or posterior probability—had to be grounded on objective (statistical "population") distributions. Thus, statements of likelihood were judged valid because, in taking a statistical parameter as given, they hypothesized the very conditions that made them "objective." But the priors for statistical parameters that Jeffreys adopted were only expressions of ignorance, in the tradition Laplace had created. They were *not* (nor were they intended as) statistical assertions about, for instance, some hyperpopulation of normal distributions from which a particular $\mathbf{N}(\mu, \sigma^2)$ was selected at random. Fisher's 1933 paper offered a new fiducial solution to the prediction problem raised by Jeffreys. To appreciate Fisher's contribution, we have to digress for a sketch of fiducial reasoning. How can there be a solution to inverse inference (from sample to population) that does not confront Bayes's theorem? How can there be posterior probability without prior probability?[8]

Consider a simplification of our heuristic example where we know the normal population variance, σ^2, only μ is not known. Fisher reasoned this way: the quantity $v = (\sqrt{n})(\bar{x} - \mu)/\sigma$ is pivotal, having a standard $\mathbf{N}(0,1)$ distribution, independent of the unknown mean μ. That is, prior to knowing $(x_1, x_2, \ldots, x_n)$, v is $\mathbf{N}(0,1)$. Fisher asserted that ignorance about μ means that after learning $(x_1, x_2, \ldots, x_n)$, still v is $\mathbf{N}(0,1)$; that is, Fisher claims that $\bar{x}$ is irrelevant to v in the absence of knowledge of μ. But given $\bar{x}$, "v is $\mathbf{N}(0,1)$" is equivalent to "μ is $\mathbf{N}(\bar{x}, \sigma^2/n)$."[9] Thus, Fisher derives a statement of inverse probability, apparently, without recourse to a specific prior for the unknown μ. It is no coincidence that the Bayesian reconstruction of this fiducial reasoning yields the same numerical conclusions based on the (improper) uniform prior, $d\mu$.

7. See Levi 1980 for discussion that the noted philosopher, C. S. Peirce, had published the outlines of confidence interval theory fifty years earlier, though it went unnoticed.

8. Modifying Savage's 1963 quip: How could Fisher make a Bayesian omelette without breaking the Bayesian eggs?

9. This reconstruction of fiducial inference as resting on an "irrelevance" step in pivotal reasoning was made clear by Jeffreys in section 7.1 of *Theory of Probability* ([1961] 1967). Hacking 1965 attempts to ground this irrelevance claim on likelihood-based reasoning. I discuss the extent to which Fisher's fiducial methods were Bayesian in Seidenfeld 1992.

In his 1933 reply to Jeffreys, Fisher solves the prediction problem in fiducial fashion, roughly as follows: the quantity $u = (s^2/\sigma^2)$ is pivotal with a χ^2 distribution (on $n - 1$ degrees of freedom). Inverting on this pivotal affords a fiducial distribution for the variance, which we may denote as $P(\sigma^2 \mid s^2)$, an inverse chi-square distribution. Given σ^2, the v pivotal affords a fiducial distribution for the unknown population mean, denoted as $P(\mu \mid \sigma^2, \bar{x})$. These correspond, exactly, to the "posterior" probabilities Jeffreys derived using Bayes's rule and his (improper) joint prior $(d\mu d\sigma/\sigma)$. In other words, Fisher was able to duplicate Jeffreys's predictive probability for a third normal variate, given the first two, without appeal to a "prior" probability, by using fiducial reasoning instead. Fiducially, given (x_1, x_2), the probability is $1/3$ that x_3 lies between the other two.

Foundations and the Appeal to Keynes's Work

Is there better evidence of a statistical dispute being foundational than that the opposing sides agree in the precise mathematical form of their answer while also disparaging the other's reasoning? Thus, by the third of the five papers, Jeffreys's rebuttal (1933) announces the need to explore what the concept "probability" means. Appealing to Bertrand Russell's synopsis of the philosophical problem of induction, Jeffreys quotes that "induction appears to me to be either disguised deduction or a mere method of making plausible guesses" (523), and, with respect to the first alternative, Jeffreys suggests three strategies: (i) induction based on "the law of contradiction"; (ii) induction based on a "law of causality"; and (iii) induction based on the theory of probability.

Of course, Jeffreys opts for the third strategy. That is, he advocates a theory of probability that relates theories and evidence. To carry this off, however, he needs (and knows that he needs) prescriptions for assigning determinate probabilities in specific cases. For that he baldly asserts, "the existence of a numerical theory of probability, however, is not enough for practical application without some rules for deciding what numbers are to be put into it. The fundamental rule is the Principle of Non-sufficient Reason, according to which propositions mutually exclusive on the same data must receive equal probabilities if there is nothing to enable us to choose between them" (1933, 528). And from here it is but a short step to confront Keynes ([1921] 1973, chap. 4) imposing objections to the Principle of Non-sufficient Reason. This he does, in terms of one of

Keynes's well-known examples. I paraphrase Keynes's objection, which Jeffreys quotes in full:

> If we are ignorant of area or populations of different countries of the world, then we should judge a man to be as likely an inhabitant of Great Britain as of France. Also, he should be judged as likely to inhabit Ireland as France, and by the same principle he should be judged equally likely to inhabit the British Isles as France. But, by additivity, the first two judgments make it twice as likely that he resides in the British Isles as in France, contradicting the third judgment.
>
> It will not do to solve this problem, asserts Keynes, by saying that because the British Isles are known to have two subdivisions (which alone tells us nothing about their relative populations), therefore, it is twice as likely for someone to reside in the British Isles as in France. ([1921] 1973, 44)

Jeffreys's reply is simple; he says that in this case Keynes neglects to relativize judgments of equipossibility to the background information of what counts as a "country." Either, argues Jeffreys, the person judges Great Britain and Ireland as separate countries or only as parts of a single country (the British Isles). There is no contradiction, argues Jeffreys, once this background assumption is fixed. That is, Jeffreys adopts Keynes's so-called "logical" interpretation of probability, where probability relates theory and statistical evidence, but he is not moved by Keynes's objections to Non-sufficient Reason.

Of course, Fisher is not satisfied with Jeffreys's reply to Keynes. Fisher agrees with Keynes's objection. There are two senses of the word *country*, and the investigator recognizes this; there is no justification for adopting one sense of country over the other for purposes of using the Principle of Non-sufficient Reason (Fisher 1934, 5).

I suggest the Jeffreys-Fisher exchange about Keynes's example, in criticism of Non-sufficient Reason, sits at the *surface* of their differences about Keynes's views on probability. There are two, more substantial, themes in Keynes's work that divide Jeffreys and Fisher:

1. Keynes argues ([1921] 1973, chap. 3) that, as a quantitative (real-valued) relation between two propositions ϕ and ψ, the "logical" probability $\mathbf{Q}(\phi \mid \psi)$ may not be defined for all pairs.[10]
2. In (the concluding) part 5 of the *Treatise*, Keynes tries to ground statistical inference on empirical premises only. Particularly for problems of inverse inference, Keynes explores ways of "inverting" the uncontroversial *direct inferences* (such as those Bernoulli's theorem provides), inferences that take us from statistical "population" to "sample" under random sampling.

10. See Kyburg's 1955 Ph.D. thesis for an important, early discussion of this theme.

These two considerations, I suggest, rather than the overt dispute over Non-sufficient Reason, are what separate Jeffreys and Fisher.

Regarding the first question, whether quantitative probability is defined for all pairs of propositions, Jeffreys argues in the affirmative.[11] By contrast, Fisher's theory admits three varieties of inductive support for solving inverse inference in the absence of prior probability for the hypothesis: significance testing, likelihood, and fiducial probability. These take increasingly restrictive background assumptions for their applicability. *Significance testing* requires a well-defined statistical null hypothesis, but (contrary to Neyman-Pearson hypothesis testing) no parametric family of statistical alternatives is supposed. *Likelihood* requires a parametrized family of statistical hypotheses. And *fiducial probability* requires, in addition, a suitable pivotal variable (or pivotal variables, in the case of several parameters). Only in the case of fiducial probability is inverse inference solved by a conclusion expressed as a (conditional) probability for the hypothesis, given the data. Thus Fisher's theory, the non-Bayesian theory, rather than Jeffreys's Bayesian theory, is closer to Keynes's position on the matter of whether (real-valued) probability is defined between all pairs of propositions.

Regarding the second point, whether inverse inference can be grounded on statistical premises alone, Fisher's fiducial probability attempts to do just that. Fiducial inference is the attempt to reduce inverse inference about a parameter to direct inference about a pivotal. By contrast, Jeffreys's Bayesian program offers a refined version of Non-sufficient Reason in which the statistical model fixes the symmetries that are used to determine the equiprobable states of "ignorance." That is, Jeffreys's solution to Keynes's objections about Non-sufficient Reason grounds the representation of ignorance on mathematical symmetries of the "chances."

For example, with a location parameter (e.g., the normal mean μ) the "prior" is uniform, and with a scale parameter (e.g., the normal variance σ^2) Jeffreys's prior is uniform in the log of the parameter. The sym-

11. This point is made explicit in Jeffreys's *Theory of Probability* ([1961] 1967), axioms 1 and 5.

metries Jeffreys uses to pick these priors are motivated by mathematical invariances in the statistical model, the empirical part of the model which relates the hypothetical parameter to the observed data through noncontroversial "direct inference." However, Fisher's fiducial solution to the same problem relies on an inversion of the statistically based "direct" probabilities for the pivotal variables. In fiducial inference, there is no appeal to mathematical symmetries in order to apply the principle of Non-sufficient Reason to form an "ignorance" prior probability. The only uses Fisher makes of Bayes's theorem require statistically based probabilities. Thus, on the second point too, I think Fisher's (non-Bayesian) theory comes closer than Jeffreys's Bayesian theory to Keynes's views on solving inverse inference by "inverting" on noncontroversial "direct" probability.

Among contemporary theories of statistical inference, H. E. Kyburg's original program of "Epistemological probability" (1974) captures both Keynesian themes. Regarding the first issue, in many common circumstances the (frequency-based) evidence is inadequate to support a real-valued Epistemological probability for a hypothesis. Then, Epistemological probability is interval-valued, rather than real-valued. With interval-valued probability, not all pairs of propositions are comparable by the simple qualitative relation ". . . is at least as probable as ——." That is, when probability goes interval-valued, it may be that neither of two propositions is at least as probable as the other—they are incomparable under this relation, just as Keynes and Fisher supposed. Second, Kyburg's Epistemological probability theory solves "inverse" inference by inverting on special relations between statistical samples and their populations that behave very much like Fisher's pivotals. Kyburg calls these "rationally representative sample" relations.

Some of the non-Bayesian aspects of Kyburg's theory are discussed (Kyburg 1977) in debates with Levi (1977) and also with me (Seidenfeld 1978). In any case, Kyburg's work on statistical inference shows how one development of the twin Keynesian themes (noted here) leads, naturally, away from the strict Bayesian position illustrated so clearly in Jeffreys's important work.

Nonparametric Inference and Non-sufficient Reason

I want to conclude by discussing how some contemporary work relates to a Keynesian theme in the Jeffreys-Fisher debate. For the specific problem

of forecasting x_3, given (x_1, x_2), when the data are i.i.d. normal and both parameters are unknown, Jeffreys's Bayesian method and Fisher's fiducial method lead to the same (numerical) results. Is there an extension beyond the normal model? Are there nonparametric versions, too?

This question is relevant because, in the spirit of the *Treatise on Probability*, if there is an extension to nonparametric fiducial inference that might identify a nonparametric Bayesian "prior," just as Jeffreys's (improper) prior serves as the Bayesian model for Fisher's fiducial inference in the case of normal data. Such a nonparametric ignorance "prior" might stand for a version of Non-sufficient Reason that is applicable *without* any particular knowledge of a statistical model.

Consider, then, the case of 3 i.i.d. real-valued data from an unknown continuous distribution F. The nonparametric version of Jeffreys's prediction problem asks whether there is an "ignorance" prior for the observables such that, given (x_1, x_2), the probability is $1/3$ that x_3 lies between them. Bruce Hill (1988) addresses the general question, for samples of size n. That is, is there a Bayesian model for nonparametric predictions where the following condition $(A_{(n)})$ holds?

> $A_{(n)}$: Given $(x_1, x_2, \ldots, x_n)$, the predictive probability is $1/(n+1)$ that x_{n+1} lies between any two (of $n-1$ many) order statistics, or lies outside either extreme value. That is:
>
> $P(x_{(i)} < x_{n+1} < x_{(i+1)} \mid x_1, x_2, \ldots, x_n) = 1/(n+1)$
> $(i = 1, \ldots, n-1)$, and
> $P(x_{n+1} < x_{(1)} \mid x_1, x_2, \ldots, x_n) = 1/(n+1)$, and
> $P(x_{n+1} > x_{(n)} \mid x_1, x_2, \ldots, x_n) = 1/(n+1)$.

Before reporting Hill's answer, note that there is a simple fiducial argument that satisfies $A_{(n)}$.[12] Let F_i be the c.d.f. for the random variable x_i. F_i is uniformly distributed on the unit interval, $F_i \sim$ U[0,1], independent of the unknown distribution F. Since the x_i are i.i.d., with common (unknown) distribution F, prior to observing $(x_1, x_2, \ldots, x_n)$, $(F_1, \ldots, F_n)$ is uniformly distributed on the n-dimensional unit-cube. That is, the F_i are independently distributed, and $(F_{(1)}, \ldots, F_{(n)})$ are just the (unobserved) order statistics from n independently distributed

12. I find the basis for this argument, ironically, in Fisher's second objection to Jeffreys (1934, 2). Hill (1988, 215) locates it, cryptically, in Fisher's 1939 remarks on "Student."

U[0,1] variates. We use these $F_{(i)}$ as pivotals, in a fiducial argument (sketched below).

Consider $A_{(2)}$, corresponding to the nonparametric version of Jeffreys's problem for predicting the third observation, given the first two. Let $\delta = F_{(2)} - F_{(1)}$, so that $0 \leq \delta \leq 1$. Note that

$$P(x_{(1)} < x_3 < x_{(2)} \mid \delta, x_1, x_2) = \delta = F_{(2)} - F_{(1)},$$

independent of the data, (x_1, x_2). If we take a fiducial step, the observed data are irrelevant to the joint distribution of $\{F_{(1)}, F_{(2)}\}$, that is, fiducially, in densities,

$$p(F_{(1)}, F_{(2)}) = p(F_{(1)}, F_{(2)} \mid x_1, x_2).$$

It is easy to verify that the density function for δ is: $p(\delta) = 2(1 - \delta)$. Then, we can write

$$\begin{aligned} &P(x_{(1)} < x_3 < x_{(2)} \mid x_1, x_2) \\ &\quad = \int_{\delta} P(x_{(1)} < x_3 < x_{(2)} \mid \delta, x_1, x_2) p(\delta \mid x_1, x_2)\, d\delta \\ &\quad = \int_0^1 \delta 2(1 - \delta)\, d\delta \\ &\quad = 1/3. \end{aligned}$$

Thus, a simple nonparametric fiducial argument leads to the prediction for the third observation, given the first two, which agrees with Jeffreys's condition for "ignorance" about the underlying (chance) distribution, F, for the observables.

The question for our inquiry is: What "ignorance" prior (over the data) duplicates this nonparametric inference? That is, relying on Fisher's fiducial inference as an acceptable solution to the nonparametric "inverse inference" (about δ), what is the corresponding Bayes model? The answer has interesting consequences for the Principle of Non-sufficient Reason.

Hill 1968 showed that, even for $A_{(1)}$ (and thus for all $A_{(n)}$, since $A_{(n)}$ entails $A_{(n-1)}$), the Bayes model cannot use a countably additive prior probability for the data. This is evidently so in Jeffreys's problem, involving $N(\mu, \sigma^2)$ data, where the improper prior density $d\mu d\sigma / \sigma$ corresponds to a finitely, but not countably additive probability.[13] Thus, Jeffreys's rule for choosing a prior to depict "ignorance," or (what amounts to the same) the Bayes model for Fisher's fiducial probability, requires non-countably additive probabilities. Hill's analysis reveals this is so also for the nonparametric version.

Apart from the mathematical point, what is urgent about the shift from countably additive to merely finitely additive probability? The following brief discussion illustrates a qualitative aspect of statistical inference that rises or falls with countable additivity. In his discussion of whether or not personal probability needs to be countably additive, de Finetti 1972 formulated the following concept of *conglomerability* of conditional probability: Let $\pi = \{h_1, \ldots, h_n, \ldots\}$ be a denumerable partition and let $E_P[\bullet]$ denote the (finitely additive) expectation with respect to probability P.

Definition: The probability P is *conglomerable* in π if, for each bounded variable X and constants k_1 and k_2, $k_1 \leq E_P[X] \leq k_2$ whenever $k_1 \leq E_P[X \mid h_i] \leq k_2$ $(i = 1, \ldots)$.[14]

About ten years ago Schervish et al. (1984) showed that conglomerability characterizes countable additivity. That is, with respect to denumerable partitions, as is evident, each countably additive probability is conglomerable in each partition; however, each finitely (but not countably) additive probability fails to be conglomerable for some event E, in some partition.[15]

For the particular case Jeffreys uses, predicting the third normal datum given the first two, based on the interesting work of Heath and Sudderth (1978), we learn that there is conglomerability in the margin of the observables (x_1, x_2) and in the margin of the two normal parameters, (μ, σ). However, in light of the Buehler-Feddersen inequality, below, we see that there is *conditional* nonconglomerability. Specifically, let $t = (x_1 + x_2)/(x_1 - x_2)$. Buehler and Feddersen (1963) established the following inequality obtains for all (μ, σ):

$$P(x_{(1)} \leq \mu \leq x_{(2)} \mid \mu, \sigma, |t| \leq 1.5) > .512.$$

Given $|t| \leq 1.5$, and applying conglomerability in (μ, σ), we obtain the

13. This is evident as the "uniform" prior $d\mu$ assigns *equal* prior probability to each unit interval of the form, $k \leq \mu < k + 1$ $(k = 0, k = \pm 1, k = \pm 2, \ldots)$. These unit intervals constitute a countable partition of the parameter space. Hence, by finite additivity, each has prior probability 0, though their countable union has prior probability 1.

14. Dubins 1975 shows that conglomerability in π is equivalent to disintegrability in π.

15. Note that the failure of conglomerability is for an event—that is, a simple random variable. Also, it depends on details in the mathematical structure of the (merely) finitely additive probability P, where the failure of conglomerability occurs can be determined by the unconditional expectations alone. This is discussed, at length, in Schervish et al. 1984.

inequality

$$P(x_{(1)} \leq \mu \leq x_{(2)} \mid |t| \leq 1.5) > .512.$$

However, by Jeffreys's (or Fisher's) analysis, the following obtains for each pair (x_1, x_2):

$$P(x_{(1)} \leq \mu \leq x_{(2)} \mid x_1, x_2) = .5;$$

hence, for pairs (x_1, x_2), which satisfy the inequality $|t| \leq 1.5$, we get:

$$P(x_{(1)} \leq \mu \leq x_{(2)} \mid x_1, x_2, |t| \leq 1.5) = .5.$$

Given $|t| \leq 1.5$, and applying conglomerability in (x_1, x_2), we obtain the contrary equality,

$$P(x_{(1)} \leq \mu \leq x_{(2)} \mid |t| \leq 1.5) = .5.$$

Thus, given $|t| \leq 1.5$, there is *conditional* nonconglomerability.[16]

One upshot of nonconglomerability is that "admissibility" fails—that is, simple dominance is not valid in denumerable partitions. So, of two statistical decisions D_1 and D_2, it may be that $E_P[D_1] < E_P[D_2]$, yet $E_P[D_1 \mid h_i] > E_P[D_2 \mid h_i]$ for each $i = 1, 2, \ldots$ That simple dominance fails raises a somewhat unusual question about the value of cost-free data. In the circumstances above, should the agent make a terminal decision between D_1 and D_2, or is it better to postpone that choice to learn, cost-free, which element of π obtains? The "prior" expectation of waiting for the new evidence and then deciding is negative!

Nonconglomerability of P thus raises a novel issue about the value of new data. Keynes ([1921] 1973, chap. 6) provides a brief but stimulating discussion about the vague notion of *weight of evidence*. For example, on the assumption that the weight of evidence for a hypothesis cannot decrease by learning something new, he shows that the precision (i.e., the inverse of the variance) of a distribution cannot index weight. That is, a conditional distribution may have larger variance. Still, it might be suggested that weight of evidence can be gauged decision theoretically, in terms of the value the new evidence provides in a sequential decision. But we see that, too, cannot serve as a universal index of weight because, for a finitely additive probability it may be that new evidence carries negative expected value; better to decide in advance of the new data!

16. See Kadane et al. 1986 for additional discussion.

If we adapt the Jeffreys-Fisher debate to a justification of improper priors in the name "ignorance," if we use that debate to try to restore the Principle of Non-sufficient Reason, then we have the following surprising price to pay: a consequence of "ignorance" is that sometimes it is decision theoretically better to remain ignorant than it is to learn!

I cannot imagine how Keynes would have accepted that. I suspect that on this score, regarding the representation of ignorance, Keynes would have placed himself outside the range of positions bracketed by Jeffreys's and by Fisher's analyses. And surely they would have each responded that that event had probability 2/3 of occurring anyway.

References

Buehler, R. J., and A. P. Feddersen. 1963. Note on a Conditional Property of Student's *t*. *Annals Math. Stat.* 34:1098–1100.

de Finetti, B. 1972. *Probability, Induction, and Statistics*. New York: Wiley.

Dubins, L. 1975. Finitely Additive Conditional Probabilities, Conglomerability and Disintegrations. *Annals of Probability* 3:89–99.

Fisher, R. A. 1933. The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters. *Proc. Royal Soc. London*, series A 139:343–48.

———. 1934. Probability Likelihood and Quantity of Information in the Logic of Uncertain Inference. *Proc. Royal Soc. London*, series A 146:1–8.

———. 1935. The Logic of Inductive Inference. *J. Roy. Stat. Soc.* 98:39–54 (with discussion, 55–82).

———. 1939. Student. *Annals of Eugenics* 9:1–9.

———. 1973. *Statistical Methods for Research Workers*. 14th ed. New York: Hafner.

Hacking, I. 1965. *Logic of Statistical Inference*. Cambridge: Cambridge University Press.

Heath, D., and W. Sudderth. 1978. On Finitely Additive Priors, Coherence, and Extended Admissibility. *Annals of Statistics* 6:333–45.

Hill, B. 1968. Posterior Distribution of Percentiles: Bayes' Theorem for Sampling from a Population. *J. Am. Stat. Assoc.* 63:677–91.

———. 1988. De Finetti's Theorem, Induction, and $A_{(n)}$ or Bayesian Nonparametric Predictive Inference. In *Bayesian Statistics*. Vol. 3. Edited by J. M. Bernardo, et al. Oxford: Oxford University Press.

Jeffreys, H. 1931. *Scientific Inference*. Cambridge: Cambridge University Press.

———. 1932. On the Theory of Errors and Least Squares. *Proc. Royal Soc. London*, series A 138:48–55.

———. 1933. Probability, Statistics, and the Theory of Errors. *Proc. Royal Soc. London*, series A 140:523–35.

———. 1934. Probability and Scientific Method. *Proc. Royal Soc. London*, series A 146:9–16.

———. [1961] 1967. *Theory of Probability*. 3d ed. Oxford: Oxford University Press.

Kadane, J. B., M. J. Schervish, and T. Seidenfeld. 1986. Statistical Implications of Finitely Additive Probability. In *Bayesian Inference and Decision Techniques*. Edited by P. K. Goel and A. Zellner. Amsterdam: Elsevier.

Keynes, John Maynard. [1921] 1973. *A Treatise on Probability*. Vol. 8 of *The Collected Writings of John Maynard Keynes*. 30 vols. Edited by D. E. Moggridge. London: Macmillan. (CW)

Kyburg, H. E. 1955. *Probability and Induction in the Cambridge School*. Ann Arbor, Mich.: University Microfilms.

———. 1974. *The Logical Foundations of Statistical Inference*. Dordrecht, Holland: Reidel.

———. 1977. Randomness and the Right Reference Class. *Journal of Philosophy* 74:501–20.

Lane, D. 1980. Fisher, Jeffreys, and the Nature of Probability. In *R. A. Fisher: An Appreciation*. Edited by S. E. Fienberg and D. V. Hinkley. New York: Springer-Verlag.

Levi, I. 1977. Direct Inference. *Journal of Philosophy* 74:5–29.

———. 1980. Induction as Self Correcting According to Peirce. In *Science, Belief and Behaviour*. Edited by D. H. Mellor. Cambridge: Cambridge University Press.

Savage, L. J. 1963. Discussion. *Bull. d'Inst. Internat. Statist.* 40:925–27.

Schervish, M. J., T. Seidenfeld, and J. B. Kadane. 1984. The Extent of Non-conglomerability of Finitely Additive Probabilities. *Zeitschrift für Wahrscheinlichkeitstheorie* 66:205–26.

Seidenfeld, T. 1978. Direct Inference and Inverse Inference. *Journal of Philosophy* 75:709–30.

———. 1992. R. A. Fisher's Fiducial Argument and Bayes' Theorem. *Statistical Science* 7:358–68.

Comment

Gregory Lilly

Fisher, Jeffreys, and Keynes

Professor Seidenfeld provides three valuable services to economists intrigued by Keynes and probability theory.

One, Seidenfeld alerts us to the Fisher-Jeffreys debate. It is always instructive (and sometimes amusing) when two giants clash over foundations, especially when, as in this case, the issues are not clouded with overly technical terminology.

Two, Seidenfeld suggests that Fisher and Jeffreys can be reference marks in an explication of Keynes's probability theory. He points out that in two important respects, Keynes is more like Fisher than Jeffreys. I was somewhat surprised at this since the traditional classification puts Jeffreys and Keynes into the Fisher-free category: theorists who tried to develop a "logical" conception of probability. Normally we think about how Jeffreys and Keynes are alike, and how Fisher and Keynes differ; new classifications tend to produce new insights—perhaps this one will, too.[1]

1. For example, Cottrell (1993, 43) has advised Keynes scholars who want to explore the connection between *A Treatise on Probability* and *The General Theory* that a focus on the idea that probability is about an objective relation between a hypothesis and an evidence statement is a misplaced focus. Instead of emphasizing the presumed objective logic of a probability-based philosophy of science, an idea that Keynes and Jeffreys share, these scholars should emphasize the idea that the probability relation may not be defined for all pairs of hypotheses and evidence statements, an idea that Keynes and Fisher share.