http://www.jstor.org

R. A. Fisher's Fiducial Argument and Bayes' Theorem

Teddy Seidenfeld

1. INTRODUCTION

In celebration of the 100th anniversary of Fisher's birth, I want to raise the subject of fiducial inference for our reflection. Shortly after Fisher's death in 1962, my teacher and friend, Henry Kyburg, addressed a conference on fiducial probability. I find it appropriate to begin with some of Kyburg's (1963) remarks:

> I am a logician and a philosopher; I have not studied statistics for very long, and so I still very quickly get out of my depth in a discussion of the technicalities of statistical inference. But I think it is important, none the less, for people whose interests lie in the area of inference as such to do the best they can in reacting to – and in having an action upon – current work in that particular kind of inference called "statistical." That this interaction is difficult for both parties is the more reason for attempting it. (p. 938)

My purpose in this essay is to try to assist that interaction by focusing on the rather vague inference pattern known as the "fiducial argument." I hope to show that, though it really is untenable as a logical argument, nonetheless, it illuminates several key foundational issues for understanding Fisher's disputes over the status of Bayes' theorem and thereby some of the continuing debates on the differences between so-called orthodox and Bayesian statistics.

Begin with the frank question: What is fiducial probability? The difficulty in answering simply is that there are too many responses to choose from. As is well known, Fisher's style was to offer heuristic examples of fiducial arguments and, too quickly, to propose (different) formal rules to generalize the examples. By contrast, among those who have attempted to reconstruct fiducial inference according to a well-expressed theory, one must single out the following five, in chronological order: Jeffreys (1961), Fraser (1961), Dempster (1963), Kyburg (1963, 1974) and Hacking (1965). So, instead of beginning with a formal definition, let us try to find out what fiducial inference is supposed to accomplish and then see whether, in fact, any answer to our question can be accepted. That is, first we shall

---

Teddy Seidenfeld is Professor in the Departments of Philosophy and Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.

determine whether the goals Fisher set for fiducial probability are even mutually consistent.

A convenient starting place for our investigation is the distinction between *direct* and *inverse* inference (or *direct* and *inverse* probability). These terms date back at least as far as Venn and appear often in those passages where Fisher attempts to explicate fiducial probability, for the reason I will give below. As a first approximation, the difference between direct and inverse probability is the difference between conditional probability for a specific (observable) event D given a statistical hypothesis S, $p(D|S)$, and conditional probability of a statistical hypothesis given the evidence of sample data, $p(S|D)$. For example, an instance of a direct probability statement is, "The probability is 0.5 of 'heads' on the next flip of this coin, given that it is a *fair* coin." Here the conditioning proposition may be understood as supplying the statistical hypothesis S that there is a binomial model (a *hypothetical population*) for flips with this coin, $\theta = 0.5$, and the event D is that the next flip lands "heads." An inverse probability statement is, "The probability is 0.4 that the coin is *fair*, given that 4 of 7 flips land 'heads'." Direct and inverse *inference*, then, denote those principles of inductive logic which determine or, at least, constrain the probability values of the direct and inverse probability statements. They would explain the probability values 0.5 and 0.4, assuming those values are inductively valid.

A maxim for direct inference which is commonplace, in so far as any inductive rule can be so described, is what Hacking (1965, p. 165) labels the *Frequency Principle*. It states loosely that, regarding the direct probability $p(D|S)$, provided *all* one knows (of relevance) about the event D is that it is an instance of the statistical "law" S, the direct conditional probability for event D is the value specified in the statistical law S. If all we know about the next flip of this coin is that it is a flip of a *fair* coin, then the direct probability is 0.5 that it lands "heads" on the next flip.

To this extent, direct probability is less problematic than inverse probability. There is no counterpart to the Frequency Principle for inverse probability. At best, $p(S|D)$ may be determined by appeal to Bayes' theorem, $p(S|D) \propto p(D|S)p(S)$: an inverse inference which involves both a direct probability $p(D|S)$ and a *prior* probability $p(S)$. It is here that Fisher sought relief from inverse inference through fiducial probabil-

ity for, as we will see in the next section, his idea was to reduce inverse to direct probability by fiducial inference. Thus, we have to explore the limits of direct inference in order to appreciate fiducial probability.

Of course, rarely do we find our background information so conveniently arranged as posited by the Frequency Principle. I may know where the coin was minted, or who flipped it last, in addition to knowing it is *fair*. Are these relevant considerations about the next flip, given the statistical hypothesis (S) that the coin is *fair*? I may know that the next flip of this coin is an instance of a different statistical law, S', as well an instance of S. It may be a flip of a coin drawn from urn U ($\theta = 0.6$), where half the coins in U are *fair* and the other half biased ($\theta = 0.7$) for "heads." By the Frequency Principle, if all I know about the next flip is that it is an instance of S', the direct probability that it lands "heads" on the next flip is 0.6, $p(D|S') = 0.6$. What if I know that the next flip is an instance of both laws S and S'? (H. Reichenbach, an advocate of a limiting frequency interpretation of probability, called this the problem of the reference class for single events.)

Invariably, these difficulties are resolved by some version of a Total Evidence principle. However, competing versions do not lead to the same results. For instance, the usual account of Neyman-Pearson statistical tests is in terms of direct probabilities of type 1 and type 2 errors. Likewise, the usual interpretation of confidence levels is as direct probabilities that the confidence interval covers the (unknown) parameter. These familiar probability claims are given in terms of direct inference based a *privileged* reference set of *repeated trials*. But how is the "repeated trial" so privileged? I. Levi argues (1980, Section 17.2) that the so-called orthodox theory involves a limitation, imposed before the trial, on the *evidence* that is available after the trial. He expresses this through a distinction between features of the data used as *evidence* (for direct probability statements), in contrast with features of the data that are used merely as *input to a statistical routine*. (A similar distinction, useful for reconciling "orthodox" and Bayesian methods, is created using I. J. Good's, 1971, *Statistician's Stooge*.) Which features of the data are to serve (post-trial) as statistical evidence for direct inference are identified (pretrial) by the *pragmatics* of the problem context, for example, by the repetitions which arise in a particular quality control setting. Only those statistical laws S valid for this distinguished set of repeated trials may be ·used. For example, as in the Buehler-Fedderson (1963) problem (cf. Section 3, below), it may be that conditional on the sample falling in a recognizable subset of the sample space, the coverage probability for a particular confidence interval is bounded away from the announced confidence level. However, because the extra information about the sam-

ple is not accepted as evidence, this analysis does not invalidate the unconditional confidence level statement. The Neyman-Pearson direct inference about the confidence level is pegged to the reference set of repeated trials through a (pretrial) constraint on which features of the statistical sample count as evidence.

Whereas the Neyman-Pearson resolution to direct inference is grounded on a *pragmatic* choice of evidential considerations, Fisher offers an *epistemological* criterion. His criterion (reported below) is a balance of what we know and what we don't know. Fisher incorporates his views about direct inference into what he calls a *semantics* for probability. The following is a succinct expression both of Fisher's dissatisfaction with the limiting frequency interpretation of probability and a sketch of his positive theory.

Remark

In the quotation below, Fisher alludes to an *aggregate* or *hypothetical population* of, for example, throws of a fair die to suggest what the binomial magnitude $\theta = 1/6$ is about; to wit, one-sixth of the hypothetical population of throws are "aces." I find Fisher's talk of *hypothetical populations* of value only as a metaphor. It facilitates the additional language of *subpopulations* and *relevant reference sets*, which are helpful in framing the real problems with direct inference. In a different setting—statistical estimation—Fisher (1973) uses the concept of a hypothetical population to justify his criterion of *Fisher Consistency* (see Seidenfeld, 1992).

> Indeed, I believe that a rather simple semantic confusion may be indicated as relevant to the issues discussed, as soon as consideration is given to the meaning that the word probability must have to anyone so much practically interested as a gambler, who, for example, stands to gain or lose money, in the event of an ace being thrown with a single die. To such a man the information supplied by a familiar mathematical statement such as:
>
> > "If a aces are thrown in n trials, the probability that the difference in absolute value between a/n and 1/6 shall exceed any positive value ϵ, however small, shall tend to zero as the number n is increased indefinitely,"
>
> will seem not merely remote, but also incomplete and lacking in definiteness in its application to the particular throw in which he is interested. Indeed, by itself it says nothing about that throw. . . . Before the limiting ratio of the whole set can be accepted as applicable to a particular throw, a second condition must be satisfied, namely that before the die is cast no such subset can be recognized. This is a necessary and sufficient condition

for the applicability of the limiting ratio of the entire aggregate of possible future throws as the probability of any one particular throw. On this condition we may think of a particular throw, or of a succession of throws, as a *random* sample from the aggregate, which is in this sense subjectively homogeneous and without recognizable stratification.

This fundamental requirement for the applicability to individual cases of the concept of classical probability shows clearly the role both of well-specified ignorance and of specific knowledge in a typical probability statement. . . . The *knowledge* required for such a statement refers to a well-defined aggregate, or population of possibilities within which the limiting frequency ratio must be exactly known. The necessary *ignorance* is specified by our inability to discriminate any of the different sub-aggregates having different frequency ratios, such as must always exist. (pp. 34–36)

We have here the ingredients for Fisher's solution to the direct inference problem. The *knowledge* he refers to is nothing more than knowledge of some statistical law S' (e.g., that the coin is chosen from the urn U) which applies to the particular event of interest, D (e.g., that the next flip lands "heads"). Then, in order to apply the statistical law S' to instance D, as in the Frequency principle, one must be ignorant of competing laws which apply to the particular D. Unfortunately, Fisher gives only rules-of-thumb, not a formal account of when laws compete. For example, suppose we know that D belongs also to a proper subpopulation of the statistical law S' and that law S (with different statistics) applies to this subpopulation. Then knowledge of S prevents D from being a random event with respect to law S'. But knowledge of S' does not prevent D from being a random event with respect to law S.

Accordingly, if you know that the next is a flip of a *fair* coin ($\theta = 0.5$), that information prevents it from being a random flip of a coin chosen from the urn U ($\theta = 0.6$), since S embeds the next flip in a relevant subpopulation with respect to the statistical law S'. Roughly put, only half of the flips from U are flips of a *fair* coin. But the situation between S and S' is asymmetric. Knowledge about the larger population pertaining to S' does not provide relevant information about the next flip once you know (S) it is a *fair* coin which is tossed.

The lack of a formal theory about "relevant" subpopulations makes Fisher's *semantics* for probability seriously incomplete. What shall we say about a case where we know that a specific event is an instance of a statistical law S^* whose hypothetical population is narrower than that for the law S we intend to use, but we don't know the statistics for this competing law?

What if we know this is a flip of a *fair* coin (S), but also it is a flip of a *fair* coin which stays up for more than 3 seconds (S^*)—and we don't know the statistics for the subpopulation of such *fair* coin flips? Does Fisher's theory prohibit the direct probability of 0.5 that this flip of the *fair* coin lands "heads" merely because we know (also) that it stays up for more than 3 seconds, though we have no statistical information about such extended flips?

There are two clear options. Fisher's program may be completed by adding clauses on direct inference which impose the onus for proof: either (A) on the challenges to the claim that D is not a random event under the statistical law S or (B) on the defense of the claim that D is a random event under the statistical law S. Under (A), merely knowing that the coin flip lasts more than 3 seconds, alone, does not prevent the direct probability based on the statistical hypothesis that it is a fair coin. That is, *unless one knows* that such extended coin flips have different statistics for landing "heads," the direct inference $p(D|S) = 0.5$ stands. Under (B), unless one *knows* that the subpopulation of extended flips of a *fair* coin has the same statistics of landing "heads" as the larger population of fair coin flips, then the extra information unseats the direct inference based on the larger population of all *fair* coin flips. As we shall see, Fisher's fiducial argument appears to rely on policy (A): additional data are not relevant to a direct inference unless you know the rival statistical law governing the subpopulation created by the added constraints. (See Kyburg's work for a systematic development of this strategy, to include the important cases where the agent has only partial [inequality] information about population statistics.)

2. A SKETCH OF THE UNIVARIATE FIDUCIAL ARGUMENT

According to Fisher, fiducial probability is special only by its genesis, not by its content. (He says this in numerous places, e.g., Fisher, 1973, p. 59.) That is, whatever we call fiducial probability must satisfy the mathematical calculus of probability, be that a countably additive or finitely additive (normed) measure. The uniqueness of fiducial probability is, supposedly, that it provides statements of inverse probability without admitting into the inference any (unwarranted) "prior" probability for statistical hypotheses, that is, without relying on Bayes' theorem to derive inverse probability from direct probability and prior probability. More accurately, fiducial inference attempts to derive inverse probability in the absence of statistically based prior probability. As Fisher (1973, p. 59) expresses it, a precondition for fiducial inference is that

there is insufficient background knowledge to determine an initial (or "prior") value for probability about unknown parameters by direct inference using, say, a hyperpopulation. By contrast, for example, a hyperpopulation is available in genetics when knowledge of the F_0 genotypes provides a direct probability "prior" for the F_1 genotype. Then, that prior may be used in Bayes' theorem with evidence of an F_2 phenotype, to determine an inverse probability about the F_1 genotypes.

In this respect, the ignorance of "priors" for fiducial inference is identical to Neyman's assumption when he proposes confidence intervals as a replacement for inverse inference. However, unlike the situation with confidence intervals, by Fisher's own claim, fiducial probability is to satisfy the same formal conditions as any other probability statement. In other words, it is fair to ask that fiducial probability satisfies the probability calculus for conditional probability, which entails there exists some Bayesian model with a "prior," whose posterior probability duplicates the fiducial probability. Of course, Fisher thought such models were mere mathematical niceties. Their prior probability is no more than a pretense to (imaginary) statistical information about a hyperpopulation for the parameter.

The following is a succinct statement of Fisher's (1973) views on the topic of such hyperpopulations for a "prior."

It should, in general, be borne in mind that the population of parametric values, having the fiducial distribution inferred from any particular sample, does not, of course, concern any population of populations from which that sampled might have been in reality chosen at random, for the evidence available concerns one population only, and tells us nothing more of any parent population that might lie behind it. Being concerned with probability, not with history, the fiducial argument, when available, shows that the information provided by the sample about this one population is logically equivalent to the information, which we might alternatively have possessed, that it had been chosen at random from an aggregate specified by the fiducial probability distribution. (pp. 124–125)

Rather than fiducial probability "being concerned with history," to use Fisher's polemical prose, his claim that fiducial probability is probability served to justify its use in Bayes' theorem for a wide range of inference problems: inference problems involving data from different experiments, problems involving nuisance parameters and problems of prediction. I return to this theme in Section 3, below.

What, then, is fiducial inference? How may one infer a statement of inverse probability without conceding Bayes' method of using a prior probability and Bayes' theorem? The short answer is as follows: reduce inverse inference to direct inference by manipulating the relevance conditions for applying the Frequency Principle to probability statements for a pivotal variable. Hacking (1965) and Kyburg (1963) made this logic eminently clear. Half a century ago, Jeffreys sketched the Bayesian model for fiducial inference. And almost 35 years ago, Lindley (1958) used that model to call Fisher (1973) on his bold assertion that fiducial probability was, after all, no different in "logical content" than "probability derived by different methods of reasoning" (p. 59). Specifically, Lindley questions whether fiducial probability is no different in logical content from inverse probability derived by Bayes' method. In the following rational reconstruction of fiducial inference, I try to clarify how close fiducial probability comes to that ideal.

Let us rehearse a familiar illustration of fiducial reasoning.

EXAMPLE 2.1. Consider the random variable x which, according to our background statistical knowledge, S, follows a Normal, $N(\mu, 1)$ distribution, but where we plead ignorance about the "unknown" mean μ. Fisher might say that, about the next observation x, we know *only* that it belongs to a "hypothetical population" of entities whose x-values are normally distributed with unit variance. Prior to observing x, given S we have no statistical basis for assigning probability to statements about μ, for example, "$p(-1 \leq \mu \leq 1)$" is undefined for Fisher. (Expressed in other words, we do not know statistics for a hyperpopulation of normal populations containing this one.) Of course, given a statistical hypothesis about μ, S_0: $\mu = 0$, direct probability about x follows by the Frequency Principle, for example, $p(-1 \leq x \leq 1|S_0) \approx 0.68$.

Define the pivotal variable $v = (x - \mu)$. It *seems* noncontroversial to claim v follows a standard Normal $N(0, 1)$ distribution, $\Phi(\cdot)$, independent of the unknown mean μ. (But see the remark, below.) Then, equally evident is direct probability about the next value v (corresponding to the next value x). That is, from the background statistical assumption, S, that $v \sim N(0, 1)$, by the Frequency Principle we assert $p(-1 \leq v \leq 1|S) \approx 0.68$.

Fiducial inference rests on the claim that the same direct probability statements obtain for the pivotal before and after observing the sample data. Fisher asserts that, for example, $p(-1 \leq v \leq 1|S) = p(-1 \leq v \leq 1|S, x) \approx 0.68$. Fiducial inference takes knowledge of the observed random variable to be irrelevant to direct inference about the pivotal variable. Then, given the datum, for example, $x = 7$, since the proposition "$-1 \leq v \leq 1$" is equivalent to the proposition "$6 \leq \mu \leq 8$," we conclude $p(6 \leq \mu \leq 8|S, x = 7) \approx 0.68$. According to fiducial inference, inverse probability about the parameter μ given datum x is reduced to direct probability about the pivotal variable v, all in the absence of a statistically based "prior" for μ.

Remark

Probability about the pivotal v is trivial for a subjective Bayesian who holds a countably additive prior over the parameter μ, $p(\mu)$. For such a person, $p(v) = \int_\mu p(v|\mu)dp(\mu) = \int_\mu \Phi(v)dp(\mu) = \Phi(v)$. However, the first equality is not guaranteed when the prior is merely a finitely additive *probability* which, unfortunately, is the case when an "improper" prior, for example, Lebesgue density, is used. But Jeffreys' Bayesian model for this fiducial inference requires the "improper" uniform prior probability. Hence, the Bayesian model for fiducial inference depends upon an additional premise: what Dubins (1975) calls "disintegrability" of probability in the partition of the parameter. As a further aside, the interesting treatment of finitely additive probability by Heath and Sudderth (1978) builds "disintegrability" into their definition of coherence. Their version of a coherent "improper" prior for μ leads to the familiar equality (above). A very different way to conclude v is distributed as a standard Normal variate uses what logicians call "universal generalization" over the statistical hypotheses for μ, relying on a theory of "chance" (see Seidenfeld, 1979, Appendix 9.1).

Are there ready-made pivotals? Generalize this univariate Normal $N(\mu, 1)$ example to continuous, univariate distributions: where, according to the background statistical knowledge, S, the random variable x has c.d.f. $F(x, \theta)$, and density $f(x, \theta) = \partial F/\partial x$ (with respect to Lebesgue measure). Then the c.d.f., itself, serves as a pivotal since, for each θ, $v_F = F(x, \theta)$ is uniformly distributed on the unit interval, $v_F \sim U[0, 1]$. By direct inference, $p(v_F|S)$ is uniform on $[0, 1]$. If fiducial inference is conducted with respect to the c.d.f. pivotal v_F, that is if $p(v_F|S) = p(v_F|S, x)$, then the induced "inverse" probability density for θ given x, $pd(\theta|S, x) = -\partial F/\partial \theta$, is advertised by Fisher (1973, p. 74) as the fiducial density for θ given x. Not every univariate setting allows fiducial inference, even when there is no "prior" for the parameter of interest. I shall defer to Section 4 a discussion about when inverting on a pivotal induces a fiducial probability. In the interim, let the following example serve as a warning that sometimes the datum, x, may be relevant to direct inference about a pivotal.

EXAMPLE 2.2. Let background statistical knowledge S assert that the random variable x has a uniform distribution on the closed interval $[0, \theta]$ for strictly positive θ, that is, the density $f(x, \theta) = 1/\theta$ for $0 < \theta$, $0 \le x \le \theta$ and $f(x, \theta) = 0$ otherwise. Then, given S, the direct probability is $1/2$ that $v_F \le 0.5$. Suppose the background information S also provides an upper bound $\theta^\#$ for θ, $\theta \le \theta^\#$. Though this information does not interfere with direct probability for the pivotal v_F, given S, knowledge of $\theta^\#$ blocks the fiducial step as $p(v_F|S) \ne p(v_F|S, x)$. This is so because given x, as a

matter of logic, the pivotal v_F has a truncated range, $0 < x/\theta^\# \le v_F \le 1$. given x, the pivotal v_F cannot be assigned a uniform probability on $[0, 1]$: logic dictates that the datum x is relevant to the c.d.f. pivotal. Hence, (simple) fiducial inference is not valid in this univariate problem.

3. FIDUCIAL PROBABILITY AND BAYES' THEOREM

Fisher's claim that fiducial probability is probability becomes the basis for its use in Bayes' theorem to solve other forms of statistical inference. Here, I illustrate three such applications: inverse inference with data of two "kinds," inverse inference involving nuisance parameters (multiparameter fiducial inference) and predictive inference.

Data of Two "Kinds"

Suppose datum x admits fiducial inference about parameter θ, but that (independent) datum y (where y's distribution also depends only on θ) does not allow fiducial inference. For instance, θ may be continuous though y is discrete and there is no acceptable pivotal connecting y and θ. (This assertion is explained in Section 4, in connection with the requirement of *smooth invertibility*.) Bayes' theorem yields: $p(\theta|x,y) \propto p(y|\theta) p(\theta|x)$. Fisher relies on fiducial inference to derive the inverse probability term "$p(\theta|x)$" and uses it in Bayes' theorem in this way, as illustrated in the next example.

EXAMPLE 3.1 (Fisher, 1973, Section 5.6). Let x be exponential $F(x, \theta) = 1 - \exp\{-x\theta\}$ for $0 < \theta$, $0 \le x$. Let y be a binomial count of a successes and b failures out of n independent trials, each trial with a chance of success $\rho = \exp\{-c\theta\}$. Then, based on datum x, there is an inverse fiducial density $pd(\theta|x) = x\exp\{-x\theta\}d\theta$. Let $\lambda = x/c$. Transformed to express inverse probability for ρ, $pd(\rho|x) = \lambda\rho^{\lambda - 1}d\rho$. By direct inference, $p(y|\rho) \propto \rho^a (1 - \rho)^b$. Hence, with the fiducial probability serving as a "prior" for ρ in Bayes' theorem, $pd(\rho|x, y) \propto \rho^{a + (\lambda - 1)} (1 - \rho)^b d\rho$.

Fiducial Inference with Nuisance Factors — the Step-by-Step Argument

Suppose, δ, the parameter of interest, is bound to a nuisance parameter ζ, $p(\text{data}|\delta, \zeta)$ depends on ζ (Fisher, 1973, Section 6.12). That is, there is no satisfactory pivotal connecting (a sufficient summary of) the data with δ alone. Instead, let the likelihood factor in two components, for example,

$$p(g, h|\delta, \zeta) = p(g|\delta, \zeta, h)p(h|\zeta),$$

where (g, h) are a jointly sufficient reduction of the data with respect to the two parameters. Suppose the second factor, $p(h|\zeta)$, supports fiducial inference to yield $p(\zeta|g, h)$. This corresponds to the claim that, in

the absence of knowledge of δ, h summarizes all the relevant evidence about ζ. (The claim makes sense, I believe, only in connection with the step-by-step method, which affords a Bayesian check for the coherence of the claim.) Last, suppose that the first factor supports fiducial inference for δ from g, given ζ and h, $p(\delta | \zeta, g, h)$. Then these terms may be combined, using Bayes' theorem, to yield

$$p(\delta | g, h) = \int_{\zeta} p(\delta | \zeta, g, h) p(\zeta | g, h) \, d\zeta.$$

This is Fisher's "step-by-step" method for solving the infamous Behrens-Fisher problem.

EXAMPLE 3.2 (Behrens-Fisher problem). Let x_i be i.i.d. $N(\mu_x, \sigma_x^2)$ $(i = 1, \ldots, n)$. Likewise, let y_i be i.i.d. $N(\mu_y, \sigma_y^2)$ $(i = 1, \ldots, n)$. All four parameters are unknown, but we are interested in the difference in means: $\delta = (\mu_x - \mu_y)$. The variances (σ_x^2, σ_y^2) are nuisance factors. Define $\zeta = \sigma_x^2 / \sigma_y^2$, the population variance ratio, and let $z = s_x^2 / s_y^2$, the sample variance ratio. Last, define the quantity

$$d' = \frac{\bar{x} - \bar{y}}{\left[\left[s_x^2 + s_y^2 \zeta \right] \left[\dfrac{\zeta + 1}{2\zeta} \right] \right]^{0.5}}.$$

Then, given ζ, there is a simple fiducial inference from d' to δ, yielding: $p(\delta | d', \zeta)$, as $p(d' | \delta, \zeta)$ is a Student's t (with $2n - 2$ d.f.), centered on the parameter of interest, δ. Fisher uses a fiducial inference from z to ζ, yielding $p(\zeta | z)$, as $p(z | \zeta)$ has Fisher's F distribution. (Here is where Fisher assumes z is sufficient for ζ in the absence of knowledge of δ.) Then these fiducial probabilities are combined, using Bayes' theorem:

$$p(\delta | \text{data}) = \int_{\zeta} p(\delta | d', \zeta) p(\zeta | z) \, d\zeta.$$

It is important to understand that there can be no "direct" fiducial argument duplicating this inference about δ (Linnik, 1963). That is, to appreciate the Bayesian aspects of the Behrens-Fisher solution, where Bayes' theorem is used to integrate out the nuisance parameter ζ, let us contrast it with the "step-by-step" fiducial method for inference about an unknown Normal mean, μ, when σ is a nuisance parameter.

EXAMPLE 3.3 (Student's t-distribution as a fiducial probability). Let x_i be i.i.d. $N(\mu, \sigma^2)$, with both parameters unknown, but with μ, alone, the parameter of interest. The two sample statistics $(\bar{x}, s^2)$ are jointly sufficient for the two parameters. Recall the likelihood for the data factors as follows:

$$p(\bar{x}, s^2 | \mu, \sigma^2) = p(\bar{x} | \mu, \sigma^2) p(s^2 | \sigma^2).$$

The second term, $p(s^2 | \sigma^2)$, supports fiducial inference

about the nuisance factor σ^2, given s^2. Fiducially, $p(\sigma^2 | s^2)$ treats σ^2 inversely proportional to a χ^2 distribution (with $n - 1$ d.f.). The first term, $p(\bar{x} | \mu, \sigma^2)$, supports fiducial inference about the parameter of interest μ, given $\bar{x}$ and σ^2. Fiducially, $p(\mu | \bar{x}, \sigma^2)$ is normal $N(\bar{x}, \sigma^2/n)$. These fiducial probabilities may be used in Bayes' theorem to solve for the marginal, inverse probability for the parameter of interest, just as in the Behrens-Fisher problem:

$$p(\mu | \bar{x}, s^2) = \int_{\sigma} p(\mu | \bar{x}, \sigma^2) p(\sigma^2 | s^2) \, d\sigma.$$

This yields the familiar Student's t-distribution $(n - 1$ d.f.) as a fiducial probability for μ. However, unlike the Behrens-Fisher distribution for δ, the t-distribution may be derived in a "direct" fiducial distribution using the pivotal variable: $t = \sqrt{n}(\mu - \bar{x})/s$, which of course has Student's t-distribution (with $n - 1$ d.f.).

Wrongly, I believe, Fisher (1973) asserts that the "direct" argument here is available as a simple, fiducial inference. He says,

It will be recognized that "Student's" distribution allows of induction of the fiducial type, for the inequality

$$\mu < (\bar{x} - ts/\sqrt{N}),$$

will be satisfied with just half the probability for which t is tabulated, if t is positive, and with the complement of this value if t is negative. The reference set for which this probability statement holds is that of the values of μ, $\bar{x}$ and s corresponding to the same sample, for all samples of a given size or all normal populations. Since $\bar{x}$ and s are jointly Sufficient for estimation, and knowledge of μ and σ a priori is absent, there is no possibility of recognizing any subset of cases, within the general set, for which any different value of the probability should hold. The unknown parameter μ has therefore a frequency distribution a posteriori defined by "Student's" distribution. (pp. 82–84)

In describing the "step-by-step" method, using Bayes' theorem to solve the joint distribution for (μ, σ) and then integrating out the nuisance parameter σ, Fisher (1973) writes,

The rigorous step-by-step demonstration of the bivariate distribution by the fiducial argument would in fact consist of the establishment of the second factor giving the distribution of σ given S, disregarding the other parameter, μ, and then of finding the first factor as the distribution of μ given $\bar{x}$ and σ. Several writers have adduced instances in which, when the formal requirements of the fiducial argument are ignored, the results of the projection of frequency elements using arti-

ficially constructed pivotal quantities may be inconsistent. When the fiducial argument itself is applicable, there can be no such inconsistency.

It will be noticed that in this simultaneous distribution μ and σ^2 are not distributed independently. Integration with respect to either variable yields the unconditional distribution of the other, and these are naturally obtainable by direct application of the fiducial argument, namely that

$$(\mu - \bar{x})/\sigma$$

is distributed as is t for $(N - 1)$ degrees of freedom, while

$$S/\sigma^2$$

is distributed as is χ^2 for $(N - 1)$ degrees of freedom. The distribution of any chosen function of μ and σ^2 can equally be obtained. (pp. 123–124)

Alas, Fisher's conclusion about the "direct" argument is unwarranted. Using the t-pivotal to shortcut the step-by-step argument does *not* produce an instance of fiducial reasoning where direct probability about the t-pivotal is unaffected by observed, sample information. True, $(\bar{x}, s)$ are jointly sufficient for (μ, σ). But that premise is logically inadequate for Fisher's conclusion about the impossibility of a recognizable reference set for the t-pivotal. There is a recognizable subpopulation having statistics that conflict with the t-distribution, as we see next.

EXAMPLE 3.3 Continued (Buehler-Feddersen problem). Let $n = 2$, so we have two (i.i.d.) observations from $N(\mu, \sigma^2)$. Trivially, there is the direct probability

$$(3.1) \qquad p(x_{\min} \le \mu \le x_{\max}|\mu, \sigma^2) = 0.5,$$

for each pair (μ, σ^2). Likewise, the fiducial "marginal" t-probability (1 d.f.) satisfies,

$$(3.2) \qquad p(x_{\min} \le \mu \le x_{\max}|x_1, x_2) = 0.5,$$

for all samples (x_1, x_2). Define the statistic $u = |x_1 + x_2| / |x_1 - x_2|$. Then, as R. J. Buehler and A. P. Feddersen proved (1963), within a year of Fisher's death,

$$(3.3) \quad p(x_{\min} \le \mu \le x_{\max} \mid \mu, \sigma^2, u \le 1.5) > 0.518,$$

for each pair (μ, σ^2). If the observed sample satisfies $(u \le 1.5)$, doesn't the inequality (3.3) point us to a subpopulation with statistics different from that for the t-distribution (on 1 d.f.), summarized in (3.2)? Given $(u \le 1.5)$, is not the fiducial step for the t-pivotal invalid? Isn't the evidence $(u \le 1.5)$ relevant to direct inference about t: $p(t) \neq p(t|u \le 1.5)$?

Let me propose a way out of this dilemma for fiducial probability, but at the expense of the "direct" argument. As Fisher asserts in the second of the conflicting quotations above, the step-by-step method is the *rigor-*

ous demonstration of fiducial inference for problems involving several parameters. That is, to avoid paradoxes from relevant reference sets, joint fiducial inference has to be related to marginal fiducial inference through Bayes' theorem. Thus, Student's t-distribution is the marginal, fiducial probability for the unknown mean μ. But that fiducial probability is not derived by a "direct" fiducial argument involving the t-pivotal.

The proposal to base fiducial inference on the step-by-step method is incomplete as a resolution of the dilemma without some explanation of Buehler and Feddersen's relevant reference-set "paradox." The answer I propose has no basis in any of Fisher's writings I am aware of. In fact, as a mathematical rather than statistical response, it is likely Fisher would have objected to it just as he vehemently objected to what he felt were excessively mathematical theories of statistics. Nonetheless, what follows is that explanation which makes the most sense to me.

The Buehler-Fedderson phenomenon does not produce a contradiction with the step-by-step fiducial argument, I propose, because fiducial inference uses a finitely and not necessarily countably additive theory of probability. Recall, by the late 1930s, Jeffreys (1961, Section 3.4, pp. 137–147) had shown that the Bayesian model for fiducial probability in two-parameter $N(\mu, \sigma^2)$ inference uses the "improper" prior density, $d\mu\, d\sigma/\sigma$. An improper prior assigns equal magnitudes of "support" to each of a countably infinite partition. For example, the uniform, Lebesgue density for a real-valued quantity μ assigns equal magnitudes to each unit interval. But there can be no countably additive probability which duplicates this feat. Only (purely) finitely additive probabilities satisfy the condition that each unit interval for μ has equal prior probability. Of course, there are different ways of representing the improper prior that mask this feature: it can be thought of as a sigma-finite measure (Renyi, 1955), or as a limit of so-called "proper" countably additive probabilities (Lindley, 1969). Regardless of how it is described, the improper prior behaves like the finitely additive probability it is! (See also the discussion by Levi, 1980, pp. 125–131, and by Kadane et al., 1986.)

The reason for emphasizing this point is that, unlike countably additive probability, finitely additive probability must admit what de Finetti called "nonconglomerability."

DEFINITION (Dubins, 1975). Say that probability p is *conglomerable* in an infinite partition $\pi = \{h_\alpha\}$ if, for each bounded random variable x,

provided $k_1 \le E_p[x \mid h_\alpha] \le k_2$ (for each element of π),
then $k_1 \le E_p[x] \le k_2$,

where "$E_p[\ ]$" denotes expectation with respect to p.

As Dubins (1975) has shown, conglomerability of p in a partition π is equivalent to p disintegrability in π.

However, Schervish, Seidenfeld and Kadane (1984) have demonstrated that each finitely additive probability that is not countably additive suffers nonconglomerability of probability (for indicator functions) in some countable partition. Hence, Bayesian models that rely on improper priors can display nonconglomerability without being inconsistent.

How does this relate to the Buehler-Feddersen "paradox"? With respect to Jeffreys' Bayesian model for fiducial inference, consider the conditional probability

$$p(\cdot \mid u \leq 1.5).$$

If this conditional probability is conglomerable in the (two-dimensional) partition of the unknown parameters, then (3.3) entails

$$p(x_{\min} \leq \mu \leq x_{\max} \mid u \leq 1.5) > 0.518.$$

If the conditional probability is conglomerable in the (two-dimensional) partition of the observed random variables, then (3.2) entails

$$p(x_{\min} \leq \mu \leq x_{\max} \mid u \leq 1.5) = 0.5.$$

Together, the conditional conglomerability assumptions make an inconsistent pair. Given $(u \leq 1.5)$, p experiences nonconglomerability in (at least) one of these two partitions. Suppose we opt to make $p(\cdot \mid u \leq 1.5)$ conglomerable in the partition of the data, as we may with Jeffreys' model. Then there is no warrant for using the relevant subset argument with the t-pivotal, based on direct (conditional) probabilities of the kind in (3.3). Moreover, under the same model, we may have $p(\cdot)$ *unconditionally* conglomerable in both partitions! (See Heath and Sudderth, 1978. Thus the event $(u \leq 1.5)$ has prior probability 0 in Jeffreys' model.) In short, there is a consistent formulation of fiducial inference which saves the step-by-step method, leading to the marginal fiducial t-distribution for μ and which, at the same time, places no weight on the existense of relevant subsets for direct inference in the form (3.3). What more than that can be asked in defense of Fisher's theory against the Buehler-Feddersen "paradox"?

Remark

To the best of my knowledge, the closest Fisher comes to recognizing the finitely additive nature of fiducial probability is in one of his discussions of the Behrens-Fisher significance test (Fisher, 1973, p. 100). There he observes that, given the null-hypothesis, the announced significance level for his test may be greater than the so-called "coverage" probability at each value of the unknown (nuisance parameter) variance ratio. Only at the limiting variance ratios (of 0 and ∞) is the coverage probability equal to the significance level, being smaller for each point in the nuisance parameter space. Hence, only with an improper prior over the nuisance parameter, which agglutinates all its mass at the two endpoints of the parameter space, can the average coverage probability equal the announced significance level.

Fiducial Prediction

Prediction offers a third variety of fiducial inference supported by using fiducial probabilities in Bayes' theorem. Suppose we observe $x_1 \sim N(\mu, 1)$ and we want to predict an (independent) $x_2 \sim N(\mu, 1)$, from the same hypothetical population. Of course, there is the "direct" pivotal argument: let $y = (x_2 - x_1)$ and $y \sim N(0, 2)$. However, if we are to incorporate these predictions with our fiducial inferences about μ, then the following use of Bayes' theorem provides the so-called "rigorous" argument. The joint likelihood factors: $p(x_1, x_2|\mu) = p(x_1|\mu)p(x_2|\mu)$. Bayes' theorem leads to the result

$$p(x_2|x_1) \propto \int_\mu p(x_2|\mu)p(\mu|x_1)\, d\mu.$$

Use the fiducial probability $p(\mu|x_1)$ in this consequence of Bayes' theorem, that is, where the conditional probability $p(\mu|x_1)$ is as $\mu \sim N(x_1, 1)$. The result is in agreement with the "direct" pivotal conclusion. Fisher (1960) gives the same analysis for a more complicated case of normal prediction when both μ and σ^2 are unknown. That problem involves the joint fiducial posterior for (μ, σ^2) given an observed sample which, then, is integrated out to yield the fiducial prediction for a second, independent sample from the same population.

4. CANONICAL PIVOTALS

When may the fiducial argument be applied to a pivotal variable? When may the observed data be irrelevant to direct inference about a pivotal? If univariate fiducial probability using the c.d.f. pivotal has a Bayesian model, that is, if $-\partial F/\partial \theta$ coheres with Bayes' theorem, Lindley (1958) answered the question. Specifically, within a fixed exponential family, combine a fiducial probability induced by datum x_1 for unknown θ with the likelihood for independent datum x_2 given θ, using Bayes' theorem, to obtain a posterior probability $p_1(\theta|x_1, x_2)$. The probability p_1 agrees with $p_2(\theta|y)$, a direct fiducial probability for θ given y (where y is a sufficient statistic for the composite data) provided the problem admits a transformation to a location parameter.

In this section, I want to suggest a Fisherian justification for a restriction on pivotal variables that insures the minimal coherence of fiducial probability afforded by Lindley's result. Assume a one-dimensional continuous, parametric model for datum x, with density $f(x|\theta)$. Let $v = g(x, \theta)$ be a pivotal, that is, a variable whose distribution is determinate without knowing the parameter θ. If knowledge of x is to be irrelevant to direct probability for v, and for this to induce a well-defined

fiducial density on θ, then (Tukey, 1957) the following three are necessary: (i) that v has the same range for each possible value of x; (ii) that v is 1-1, with single valued inverse; and (iii) this inverse is continuous with continuous derivatives. Tukey calls a pivotal *smoothly invertible* when it satisfies these requirements. Smooth invertibility is equivalent to conditions that Fisher (1973, pp. 73–74 and p. 179) called for. Evidently, Example 2.2 involves a failure of clause (i), which helps to explain why the fiducial argument does not apply there.

Smooth invertibility of a pivotal does not suffice for coherence of fiducial inference, however, as illustrated by Lindley's (1958, p. 229) counterexample or Good's (1965, Appendix A). With motivation to follow in due course, define pivotal v to be *canonical* provided it is smoothly invertible and its distribution $p(v)$ is the same as the conditional distribution $p(x|\theta^*)$ of the observable x for some value of the unknown parameter $\theta = \theta^*$.

In advance of proposing a justification for this added constraint on pivotal variables, examine the following illustrations of canonical pivotals for several of the fiducial inferences used elsewhere in this essay. In each case, Tukey's condition of smooth invertibility is easily verified.

EXAMPLE 4.1 (Normal mean). Let $x \sim N(\mu, 1)$. Define pivotal $v = (x - \mu)$, so $v \sim N(0, 1)$. That is, v is canonical for $\mu^* = 0$.

EXAMPLE 4.2 (Exponential distribution). Let c.d.f. $F(x, \theta) = 1 - \exp\{-x\theta\}$ for positive x and θ. Define pivotal $v = x\theta$, so v is exponential with parameter 1. That is, v is canonical for $\theta^* = 1$.

EXAMPLE 4.3 (Uniform distribution). Let c.d.f. $F(x, q) = x/\theta$ for positive θ and $0 \leq x \leq \theta$. Let $v = F$, the c.d.f. itself. Then v is canonical for $\theta^* = 1$. Note that, with an appeal to the step-by-step method, this pivotal provides a fiducial solution to the problem of Example 2.2. Treat the information about the upper bound $\theta^\#$ as evidence for conditioning. Then, $p[\theta|x, (\theta \leq \theta^\#)]$ is simply the conditional probability derived from the fiducial probability $p(\theta|x)$.

EXAMPLE 4.4 (Normal variance). Suppose s^2 is the sample variance of n, i.i.d. normal $N(\mu, \sigma^2)$ observations. Let the pivotal $v = s^2/\sigma^2$. Then v has a χ^2 distribution (with $n - 1$ d.f.), and v is canonical for $\sigma^* = 1$.

Remark

The fiducial argument associated with Example 4.4 is one of the steps used in the step-by-step fiducial inference for the two-parameter Normal problem of Example 3.3. This "step" raises again the knotty question of how to justify the claim that s^2 is sufficient for σ^2 in the absence of prior knowledge of μ. A similar problem was pointed out in connection with example

3.2. Note, too, that if μ is known the pivotal for fiducial inference about σ^2 is not s^2/σ^2 since, then, s^2 is not a sufficient statistic of the data for σ^2. Instead, if μ is known, the fiducial inference uses the (sufficient) sample sum-of-squares around μ.

There are two justifications for the condition that pivotals be *canonical*, a Bayesian reason and a Fisherian reason. First, regarding Bayesian models for fiducial inference, in the setting considered by Lindley canonical pivotals are coherent. That is, with canonical pivotals in the exponential family, the inference problem can be transformed to a location parameter. (See Seidenfeld, 1979, Appendix 9.3.) Second, fiducial inference with canonical pivotals supplies a link between (Fisherian) tests of significance and fiducial probability.

As early as 1930, in his first attempt at fiducial inference involving the correlation of a bivariate normal population, Fisher proposed "fiducial intervals" as sets of unrejected null-hypotheses (Fisher, 1930, p. 533). Significance tests remain as popular in applied statistics as they are enigmatic to Bayesian inference. Nonetheless, since fiducial inference with canonical pivotals reduces to inference for a location parameter, there is a simple tie among fiducial probability, likelihood and significance levels.

Suppose, in location form, significance levels are determined by a so-called probabilistic discrepancy ranking—to use the language of H. Cramér (1946). That is, call a sample outcome o_1 *more discrepant with the* (null) *hypothesis* $H_{\theta 0}$ *than* sample outcome o_2 if and only if $p(o_2|\theta_0) < p(o_1|\theta_0)$. This ordering makes outcomes "rarer" under an hypothesis inversely proportional to their probability given that hypothesis. (Recall Fisher's [1973] interpretation of a significance test as concluding, "*Either* an exceptionally rare chance has occurred, *or* the [null hypothesis] theory of random distribution is not true," p. 42.) It corresponds to Fisher's (1973, Section 4.4) discrepancy ranking for his "exact" test for independence in contingency tables. Likewise, identify the *significance level* of an outcome o, for an hypothesis H, as the (conditional) probability of the set of outcomes as discrepant as o given H.

We may use the probabilistic discrepancy ranking also to predict outcomes of the pivotal variable. That is, with probability 0.95, we predict there will not be a pivotal outcome with significance level as low as 0.05. If we retain this "fiducial prediction" in the face of the observed datum, we create a set of parameter values with fiducial probability 0.95. These parameter values correspond to just those "null" values for which the observed datum is not discrepant at the 0.05 significance level (or less). Also, since the prior for the Bayesian model of fiducial inference is uniform (when in location parameter form), this set is the 0.95 fiducial probability set of "most likely" hypotheses, given the datum observed.

EXAMPLE 4.5 (Normal, fiducial intervals). Let $x \sim$ N$(\mu, 1)$, with canonical pivotal $v = (x - \mu)$, and so $v \sim$ N$(0, 1)$. Under the probabilistic discrepancy ranking for outcomes, the 0.05 significance level for hypothesis $H: \mu = k$, is (approximately) the set $\{x: |x - k| > 2\}$. Likewise, a prediction about v, using the same index of discrepancy, is $p\{-2 \leq v \leq 2\} \approx 0.95$. The fiducial prediction about v, then, is $p\{-2 \leq v \leq 2 \mid x\} \approx 0.95$. But this fiducial prediction yields a 0.95 "fiducial interval" of μ values coinciding with the set of values $\{\mu: x - 2 \leq \mu \leq x + 2\}$ of hypotheses which are not significant at the 0.05 level (or less). As is well known, this interval also results from a likelihood ratio test. In fact, this agreement between fiducial intervals and significance tests requires, only, that "fiducial prediction" for the canonical pivotal be based on the same discrepancy index as is used to determine significance levels for the statistical hypotheses. In particular, the probabilistic discrepancy index is useful, in addition, for establishing the tie to likelihood ratio tests for location parameters.

5. NON-BAYESIAN ASPECTS OF FIDUCIAL INFERENCE AND CONCLUSION

The fiducial argument displays its non-Bayesian character through reliance on the sample space of possible observations to locate its Bayesian model. That is, the prior for fiducial inference may depend upon which component of the likelihood is used to drive the fiducial argument.

EXAMPLE 5.1 (Inconsistent fiducial inferences using Bayes' theorem). Let $x \sim$ N$(\mu, 1)$ and, independently, let $y \sim$ N$(v, 1)$, where $\mu = v^3$. Such variety of data might arise by using different measurement techniques for the same (theoretical) unknown parameter. However, because μ and v are not linearly related, there is no real-valued sufficient statistic for the pair (x, y) — they are minimally sufficient by themselves — and Lindley's result does not apply.

The joint likelihood factors are as follows:

$$p(x, y|\mu) = p(x|\mu) = p(x|\mu)p(y|\mu),$$

so there is the opportunity for using Bayes theorem with a fiducial probability based on (either) one of these factors:

$$p(\mu|x, y) \propto p(x|\mu)p(\mu|y)$$

and

$$p(\mu|x, y) \propto p(y|\mu)p(\mu|x).$$

However, contrary to Bayes' theorem, the inverse fiducial probability, $p(\mu|x, y)$, depends upon which factor of the likelihood is used for fiducial inference. This is

readily understood in terms of Jeffreys' Bayesian model for the fiducial argument. When we create $p(\mu|y)$ by fiducial reasoning, we use the canonical pivotal $(y - v)$ whose Bayesian model requires a uniform ("improper") prior over v. When, instead, we create $p(\mu|x)$ by fiducial reasoning, we use the canonical pivotal $(x - \mu)$ whose Bayesian model requires a uniform ("improper") prior over μ. Because μ and v are nonlinear transformations of the same quantity, it is impossible to have a uniform distribution simultaneously over both.

This fault in fiducial inference, then, is yet another version of a very old problem with Laplace's *Principle of Insufficient Reason*. If "ignorance" over a set of possibilities is to be represented by a uniform probability, to capture the symmetry which "ignorance" entails, then we get mutually inconsistent representations of the same state of "ignorance" merely through an equivalent reparameterization of the set of possibilities. Both μ and v parameterize the same set of possibilities. Through the lens of Jeffreys' Bayesian model, we see where the fiducial step leads Fisher's assumption of prior ignorance into the paradoxes of *Insufficient Reason*. That is to say, making the data irrelevant to direct probability for the pivotal conflicts with Fisher's claim that there is *no* prior probability over the parameter. The contradiction is not that *some* "prior" is required for a Bayesian model of fiducial inference. Rather, it is using fiducial probability in Bayes' theorem that leads to a contradiction about which "prior" represents the same state of ignorance.

To be fair to Fisher, this problem is not his alone. Example 5.1 is a challenging exercise for a wide variety of (what Savage called) "necessitarian" theories: theories that try to find privileged distributions to represent "ignorance." The Example 5.1 applies to Jeffreys' (1961, Section 3.10) theory of Invariance, which uses Information theory to fix symmetries preserved in a prior. It applies to Fraser's (1968) group theoretic, Structural Inference. And it applies to Jaynes' (1983) program of Maximum Entropy.

In his 1957 paper, "The Underworld of Probability," Fisher proposes a modified fiducial argument with inequalities in place of equalities of probabilities, for example, fiducial conclusions of the form $p(\theta \geq 0) > 0.5$ to replace statements like, $p(\theta \geq 0) = 0.5$. This idea relates to current research using sets of probabilities, rather than a single probability, to represent an inductive conclusion. Can ignorance be depicted by a large set of prior probabilities? Explicit connection of this approach with fiducial inference is found in A. P. Dempster's (1966) work and in H. E. Kyburg's (1961, 1974) novel theory. Perhaps it is premature to say we have seen the end of the fiducial idea?!

In the conference on fiducial probability (from which I quoted Kyburg's remarks to begin this paper), Savage (1963) wrote:

The aim of fiducial probability . . . seems to be what I term "making the Bayesian omelette without breaking the Bayesian eggs." (p. 926)

In that sense, fiducial probability is impossible. You cannot reduce inverse to direct inference. As with many great intellectual contributions, what is of lasting value is what we learn trying to understand Fisher's insights on fiducial probability. His solution to the Behrens-Fisher problem, for example, was a brilliant treatment of nuisance parameters using Bayes' theorem. His instincts about "recognizable subpopulations" led to work on so-called relevant reference sets, a subject of continuing research even from the "orthodox" point of view. And, exploration of multiparameter fiducial inference helped to expose puzzles of improper priors — an area still ripe with controversy (Seidenfeld, 1982). In this sense, "the fiducial argument is 'learning from Fisher'" (Savage, 1963, p. 926). So interpreted, it certainly remains a valuable addition to the statistical lore.

ACKNOWLEDGMENTS

REFERENCES

BUEHLER, R. J. and FEDDERSEN, A. P. (1963). Note on a conditional property of Student's t. *Ann. Math. Statist.* 34 1098-1100.

CRAMÉR, H. C. (1946). *Mathematical Methods of Statistics.* Report 1971, Princeton Univ. Press.

DEMPSTER, A. P. (1963). On direct probabilities. *J. Roy. Statist. Soc. Ser. B* 35 100-110.

DEMPSTER, A. P. (1966). New methods for reasoning towards posterior distributions based on sample data. *Ann. Math. Statist.* 37 355-374.

DUBINS, L. (1975). Finitely additive conditional probabilities, conglomerability and disintegrations. *Ann. Probab.* 3 89-99.

FISHER, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society* 26 528-535.

FISHER, R. A. (1957). The underworld of probability. *Sankhyā* 18 201-210.

FISHER, R. A. (1960). On some extensions of Bayesian inference proposed by Mr. Lindley. *J. Roy. Statist. Soc. Ser. B* 22 299-301.

FISHER, R. A. (1973). *Statistical Methods and Scientific Inference,* 3rd ed. Hafner, New York.

FRASER, D. A. S. (1961). The fiducial method and invariance. *Biometrika* 48 261-280.

FRASER, D. A. S. (1968). *The Structure of Inference.* Wiley, New York.

GOOD, I. J. (1965). *The Estimation of Probabilities.* MIT Press.

GOOD, I. J. (1971). Twenty seven principles of rationality. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) 124-127. Holt, Rinehart, and Winston, Toronto.

HACKING, I. (1965). *Logic of Statistical Inference.* Cambridge Univ. Press.

HEATH, J. D. and SUDDERTH, W. (1978). On finitely additive priors, coherence, and extended admissibility. *Ann. Statist.* 6 333-345.

JAYNES, E. T. (1983). *Papers on Probability, Statistics and Statistical Physics* (R. Rosenkrantz, ed.). Reidel, Dordrecht.

JEFFREYS, H. (1961). *Theory of Probability,* 3rd ed. Oxford Univ. Press.

KADANE, J. B., SCHERVISH, M. J. and SEIDENFELD, T. (1986). Statistical implications of finitely additive probability. In *Bayesian Inference and Decision Techniques* (P. Goel and A. Zellner, eds.) 59-76. Elsevier, New York.

KYBURG, H. E. (1961). *Probability and Logic of Rational Belief.* Wesleyan Univ. Press.

KYBURG, H. E. (1963). Logical and fiducial probability. *Bull. Inst. Internat. Statist.* 40 884-901, 938-939.

KYBURG, H. E. (1974). *Logical Foundations of Statistical Inference.* Reidel, Boston.

LEVI, I. (1980). *The Enterprise of Knowledge.* MIT Press.

LINDLEY, D. V. (1958). Fiducial distributions and Bayes' theorem. *J. Roy. Statist. Soc. Ser. B* 20 102-107.

LINDLEY, D. V. (1969). *Introduction to Probability and Statistics from a Bayesian Viewpoint.* Cambridge Univ. Press.

LINNIK, Y. V. (1963). On the Behrens-Fisher problem. *Bull. Inst. Internat. Statist.* 40 833-841.

RENYI, A. (1955). On a new axiomatic theory of probability. *Acta Math. Hungar.* 6 285-335.

SAVAGE, L. J. (1963). Discussion. *Bull. Inst. Internat. Statist.* 40 925-927.

SCHERVISH, M. J., SEIDENFELD, T. and KADANE J. B. (1984). The extent of non-conglomerability of finitely additive probabilities. *Z. Wahrsh. Verw. Gebiete* 66 205-226.

SEIDENFELD, T. (1979). *Philosophical Problems of Statistical Inference.* Reidel, Dordrecht.

SEIDENFELD, T. (1982). Paradoxes of conglomerability and fiducial inference. In *Proceedings of the 6th International Congress on Logic, Methodology and Philosophy of Science* (J. Los and H. Pfeiffer, eds.) 395-412. North Holland, Amsterdam.

SEIDENFELD, T. (1992). R. A. Fisher on the design of experiments and statistical estimation. In *The Founders of Evolutionary Genetics* (S. Sarkav, ed.). Kluwer, Dordrecht. In press.

TUKEY, J. W. (1957). Some examples with fiducial relevance. *Ann. Math. Statist.* 28 687-695.