International Statistical Review (2005), **73,** 2, 289–307, Printed in Wales by Cambrian Printers

The Statistical Education of Harold Jeffreys

John Aldrich

Economics Division, School of Social Sciences, University of Southampton, Southampton, SO17 1BJ, UK. E-mail: john.aldrich@soton.ac.uk

Summary

The paper considers the statistical work of the physicist Harold Jeffreys. In 1933–4 Jeffreys had a controversy with R.A. Fisher, the leading statistician of the time. Prior to the encounter, Jeffreys had worked on probability as the basis for scientific inference and had used methods from the theory of errors in astronomy and seismology. He had also started to rework the theory of errors on the basis of his theory of probability. After the encounter Jeffreys produced a full-scale Bayesian treatment of statistics in the form of his *Theory of Probability.*

Key words: Harold Jeffreys; R.A. Fisher; Bayesian inference; Fiducial argument.

Introduction

In the course of the 1930s Harold Jeffreys, applied mathematician, physicist and philosopher of science, became a statistician. Or at least, he produced a major work on statistics, the *Theory of Probability* (1939), and joined in the domestic quarrels of statisticians on such matters as randomisation. How he came to this new career is the subject of this paper.

The *Probability* appeared in a series of monographs on physics but R.A. Fisher provided the statistical point of reference. Jeffreys admired Fisher, the pre-eminent "modern statistician", but disagreed with him totally on the fundamental issues of the interpretation of probability and the validity of the Bayesian ("inverse") argument. Jeffreys's statistical education began rather publicly in 1933–1934 with a controversy with Fisher. Yet, looking back in the *Probability* Jeffreys (p. vi) wrote, "I must offer my warmest thanks to Professor R.A. Fisher ... for [his] kindness in answering numerous questions from a not very docile pupil."

Section 1 describes Jeffreys before the encounter, developing probability as the basis for scientific inference, using the theory of errors in his astronomical and seismological research and finally reworking the theory of errors. Sections 2 and 3 treat the controversy with Fisher and its effect on Jeffreys. Section 4 examines what the *Probability* did with "modern" statistics, while Section 5 examines what the statisticians made of the *Probability*.

1 Probability, Seismology and Errors

Harold Jeffreys (1891–1989) went to Cambridge University as an undergraduate in 1910 and stayed, applying mathematics to celestial mechanics, fluid dynamics, meteorology, geophysics and probability. Cambridge influenced both the general direction of his research and the details. Jeffreys learnt how to apply mathematics from George Darwin's writings, while James Jeans and Arthur Eddington were present influences. The *Principia Mathematica* of Whitehead & Russell (1910–1913) gave Jeffreys a model for his work on the foundations of probability. Howie (2002) provides a rich account of Jeffreys's personal and intellectual life, especially in the 1920s. Good (1980) and Geisser (1980) give brief accounts of Jeffreys's work in probability. There is a small annotated bibliography and sketch of Jeffreys's career in Aldrich (2003a).

The statistical project of the 1930s developed from two of Jeffreys's long-standing interests—in founding the principles of scientific inference on probability and in using the theory of errors in empirical astronomy and seismology.

1.1 Probability and Scientific Inference

Jeffreys (1963, p. 407) described how his work in geophysics and cosmogony stimulated an interest in the principles of scientific inquiry: both subjects involved huge extrapolations from observations on a laboratory scale and a "usual criticism was that such extrapolation was speculative and unverifiable." He recalled reading Karl Pearson's *Grammar of Science* (1911) in 1914 and what had impressed him: "Pearson's attitude was that the laws are not established with certainty but can have a high degree of probability on the data, and he outlined a theory of how this can happen." Jeffreys's task in probability was to improve on Pearson's theory.

Yet Jeffreys's early publications with Dorothy Wrinch (1894–1976) do not reflect Pearson so much as W.E. Johnson, with whom Wrinch had studied. Johnson had a curious position in Cambridge probability. Not only did the followers, who included C.D. Broad and J.M. Keynes, publish before the master but, as Zabell (1982) shows, there were interesting ideas in Johnson's writings, especially his (1932), on such topics as exchangeability, which the followers, including Wrinch and Jeffreys, did *not* follow up. The only Cambridge publication mentioned by Wrinch and Jeffreys is Broad's "The relation between induction and probability" (1918).

Jeffreys also recalled how he was reacting against the philosophical views of other physicists. In 1937 he told Fisher, "it was my disagreement with Eddington [over whether science could be generated from pure thought] that first brought me into the subject in 1919." (Bennett, 1990, p. 162). However the first Wrinch–Jeffreys paper, "On some aspects of the theory of probability", mentions relativity but not Eddington's position on the matter. It presents alternative views of probability—classical, frequentist and logical, they would be called today—and plumps for the "undefined concept" view, i.e. the logical view.

By extending classical logic to cases where premisses do not establish conclusions with absolute certainty, Wrinch & Jeffreys (1919, p. 716) hoped to find a rational basis for what scientists *do*. Venn's (1888) notion of probability as a limiting frequency could not achieve this:

> one could attach no meaning to a statement that it is probable that the solar system was formed by the disruptive approach of a star larger than the sun . . . . Yet such cases as these are the very ones where the notion of probability is particularly valuable in science, and any definition that will not cover them is not satisfactory.

Jeffreys never accepted that the limit made sense either in the Venn form or in the later form due to von Mises. Moreover, Venn's (1888, p. 731) probability failed in a crucial respect:

> The "undefined concept" view of probability can be developed so as to yield a theory of induction adequate for scientific purposes. There are difficulties in the way of obtaining such a theory from the frequency view, and we conclude that the balance is in favour of the "undefined concept" view.

The conclusion underlines the commitment of Wrinch and Jeffreys to the principle that induction is based on probability.

Wrinch & Jeffreys (1919, p. 722) defend one axiom with the appeal, "If then we wish to retain the customary applications of the theory (and this seems desirable at any cost) we must assume that axiom 2 is correct." The axiom states:

> If in one combination [of proposition and data] the proposition is more probable relative to the data than in another, the number corresponding to the first is greater than that corresponding to the second.

It "yields as an obvious corollary the famous 'principle of sufficient reason'; according to this, equal probabilities are assigned to propositions relative to data when the data give no reason for expecting any one rather than any other."

Keynes's *Treatise on Probability* (1921) has a chapter on "the paradoxical and even contradictory conclusions" to which the principle leads. Keynes (p. 48) reviewed such objections as the von Kries paradox whereby a uniform distribution applied to the specific density is inconsistent with a uniform distribution applied to its reciprocal, the specific volume. Wrinch and Jeffreys were less troubled and Jeffreys (1922, p. 132) remarked that Keynes "criticises severely many previous applications of this principle (so severely that an unprepared reader is likely to be betrayed into expect him to reject the principle altogether)."

Wrinch & Jeffreys (1919, p. 725) put their probability machine to work on a problem of Broad's, involving the kind of probability inference for which "detailed treatment is most applicable":

> Suppose that a bag contains m balls, an unknown number of which are white. Of these $p+q$ have been drawn and not replaced; p of them have been white and q not white. What is the probability that the number of white balls in the bag is n?

This "problem of sampling induction" was the paradigm scientific inference problem, as it had been for Jevons and Pearson. It was also a problem in statistical inference, though statisticians favoured the variant, where the number of balls is infinite and the proportion of white balls is to be determined. Wrinch and Jeffreys seem unaware of their efforts.

Wrinch and Jeffreys review the results for the case when $f(n)$, the prior probability of any particular number of white balls, is "independent of n". They consider such a uniform prior unrealistic but do not examine any alternatives for their point is, "that when the sample is large enough the prior probabilities of different constitutions of the whole do not usually affect appreciably the probabilities inferred after the samples have been taken." (p. 731).

Broad had obtained a new and striking result for the uniform prior scheme: if q is 0, the probability that *all* the balls in the bag are white is $(p+1)/(m+1)$. Thus if "enumerative induction" is to produce a large probability that a generalisation is true, the proportion of instances examined must be large. In 1980 Jeffreys (p. 452) recalled that this "preposterous" conclusion had set him and Wrinch off. In 1919, however, Wrinch and Jeffreys *agreed* with Broad, "no such [universal] law can derive a reasonable probability from experience alone; some further datum is required" and they illustrated what this datum might be: "if we consider that either Einstein's or Silberstein's form of the principle of general relativity is true, a single fact contradictory to one would amount to a proof of the other".

The second Wrinch & Jeffreys paper on probability, "On certain fundamental principles of scientific inquiry", (in two parts, 1921/23) confronted the following dilemma (1923, p. 368):

> It is a universal belief among physicists that when a sufficient number of inferences from a quantitative law have been verified, the probability of the correctness of the next inference from it may be made to approach indefinitely near to unity. [T]his proposition is not easily reconcilable with the other proposition, also believed by some physicists, that all laws of infinite class are equally probable *à priori*.

Surprisingly, given their former endorsement of Broad's conclusion, Wrinch and Jeffreys accept the physicists' belief and resolve the dilemma by giving up the second proposition, arguing that the set of possible laws is denumerable and that the laws can be arranged in order of decreasing *simplicity* with the probabilities of these laws forming a convergent series with sum unity. The simplicity principle became a permanent feature of Jeffreys's system.

Jeffreys appears to have used the probability formalism only once in his scientific writing in the 1920s. In *The Earth* he (1924, p. 259) makes the case for the tidal theory of the earth's formation and the conclusion—translated into ordinary language—is:

> given the empirical data of physics and the existence of the solar system as it is, it is practically certain both that the laws of physics used in the argument are true and that the initial conditions required for the tidal theory once occurred; and therefore that the tidal theory is true.

While the Wrinch–Jeffreys project identified the probability *implicit* in the practice of science, it did not touch the probability already *explicit*—e.g. in gas theory, or in the combination of observations. That was to change, and the *Probability* would be one result.

1.2 Least Squares & Observational Seismology

The theory of errors was part of the outfit of the empirical researcher. In Cambridge the authorities on the subject were the astronomers and Eddington lectured to Jeffreys in 1913—see Howie (2002, p. 162). Jeffreys's early use of the theory of errors was not especially innovative. In an astronomical investigation testing Einstein's theory he (1916) used weighted least squares to find the unknowns in a small number of dynamic equations, a technique going back to Gauss. His move into seismology—which started with Wrinch & Jeffreys (1923)—did not require a new outfit. Least squares was still the main technique, although there were some special features, e.g. rougher data and sometimes many more normal equations to solve.

In the 1920s Jeffreys used statistical methods without reflecting much upon them. His main source for *theory* was Whittaker & Robinson's *Calculus of Observations* (1924). This was a wonderful compendium of ideas and methods but it had no single viewpoint: e.g. on pp. 215–226 Gauss's Bayesian and non-Bayesian "proofs" of least squares just follow one another. It covered some biometric work but this was rather old; there was no mention of the very recent work of Fisher on statistical inference and least squares/regression.

1.3 The Theory of Errors: Needing "Some Modification"

Jeffreys's book *Scientific Inference* (1931a) returned to the Wrinch and Jeffreys topics, geometry, dynamics and probability. There was a solution to the Broad puzzle in a "non-quantitative simplicity postulate" (pp. 191–197): "$f(r)$ ... should be taken to be inversely proportional to $r(n-r)$." A chapter on "errors" included a response to Jeffreys's sense that "the usual presentation of the theory of errors needed some modification" (1932b, p. 48).

The principal modification (1931a, pp. 66–70) involved inference to the true value, x, when the errors follow the "normal law" with the precision *unknown*. In fact, the "usual presentation" was repeatedly extended to accommodate unknown precision. Pfanzagl & Sheynin (1996) review these efforts. Jeffreys (1931a, p. 66) adopted the usual prior for x: "In most cases [this] is nearly uniformly distributed, at any rate over a range several times that covered by the observations." However the position is "different with regard to h". (Jeffreys, like other error theorists, used the quantity $h = \frac{1}{\sigma\sqrt{2}}$.) Jeffreys treated the parameters x and h separately, taking it for granted that they are prior independent. The novel element in his treatment was the recognition that the argument *might* be based on a second measure, σ the "standard error" as he called it.

> Initially we may have no special views about the probability of one value of h rather than another, but we do at least know that negative values are excluded, since they would

> imply negative probabilities. Again, x is not usually in fact a number; it is usually a length or an interval of time, and h is a reciprocal of whatever kind of magnitude x is, while the standard error σ is the same kind of quantity as x. There seems to be no special reason for measuring the precision in terms of h rather than σ, and their product is constant, so that
>
> $$d \log h + d \log \sigma = 0.$$
>
> If then $P(x, h)dh$ is proportional to dh/h or $d\sigma/\sigma$, an ambiguity is removed.

Jeffreys's attention to invariance was new. Other Bayesian "modifiers" took the joint prior as uniform—in whatever parameterisation was adopted—without discussion. Jeffreys (1931a, p. 69) derives the "probability of a value of x in the range dx, irrespective of h".

"On the Theory of Errors and Least Squares" (Jeffreys, 1932b) extended the inverse argument to normal regression and put the case for dh/h in an "alternative form" conforming to a second way of "testing" a prior. Instead of asking what distribution "describes our *a priori* knowledge", one asks what distribution "is consistent with facts otherwise known about the posterior probability on certain types of data." Jeffreys (1933b, p. 530).

Jeffreys (1932b, p. 48) showed that dh/h is the only prior that generates a predictive density embodying the following "fact":

> Two measures are made: what is the probability that the third observation will lie between them? The answer is easily seen to be one-third. For the law says nothing about the order of occurrence of errors of different amounts, and therefore the middle one is equally likely to be the first, second, or third made (provided, of course, that we know nothing about the probable range of error already).

Fisher (1933, p. 344) summarised the argument before attacking it (see §2.2 below):

> (a) the probability of the first two observations having assigned values is expressed in terms of the two parameters . . . of the population; (b) introducing the probability *a priori* of the two parameters having assigned values, their posterior probability of having them is obtained; (c) the probability of the third observation is found and integrated over all possible values of the parameters; (d) the expression so obtained is equated to $1/3$ . . .

Jeffreys did not linger over this analysis for his seismological research (1931b) was presenting bigger problems. "An alternative to the rejection of observations" (1932a, p. 78) addressed one, "The time of arrival of a given wave may normally have a standard error of order 7 seconds . . . but about a fifth of the observations are affected by an uncertainty of the order of 20 seconds." Jeffreys treated the observations as generated from a mixture of normals and then combined his analyses of the normal mean and sampling induction problems.

His programme of developing methods for reducing seismological observations was disturbed by the encounter with statistics. This could not have held much promise. He (1931a, pp. 33–34) thought statistics was about large samples: "We do not evaluate the prior probability in practical sampling because we do not need to; we swamp it automatically when we take a sufficiently large sample. It is this principle that constitutes the theoretical justification of statistical methods." The "statistical theory of probability" was what he called Venn's theory.

2 The Encounter with Fisher

At the time Jeffreys was learning the theory of errors the most dynamic person doing what we *now* call *statistics* was Karl Pearson. Pearson had started in the same intellectual community as George

Darwin but his biometry had moved a long way from physics. 'Statistics' in the dominant sense of economic and social statistics was even more remote. Yule was in Jeffreys's college from 1913 but they seem to have only started talking statistics in the '30s.

The arrival on the scene of R.A. Fisher (1890–1962) upset the established categories and Jeffreys's "modern statisticians" reflected the new Fisherian reality. Like Jeffreys, Fisher was taught by Cambridge astronomers but his interests in agricultural statistics and genetics had taken him out of Jeffreys's way. Before 1933 each noticed the other's work *once* and dismissed it. Had they looked closer, each would have found the other caught in a fallacy *he* had exploded. Eventually, however, Jeffreys decided that Fisher was worth some attention. For the very large literature on Fisher's work see the bibliography in Aldrich (2003b).

2.1 The Practical Worker and the Mathematician

Jeffreys (1922) welcomed Keynes's *Treatise on Probability* and only criticised its "excessive caution". In *his* review Fisher (1922/3, p. 46) commented on Jeffreys

> Dr. Jeffries [sic] gives such an indulgent account as might be expected from one who has recently interested himself in some logical aspects of the subject, on somewhat similar, albeit sounder and more tolerant lines. To the practical worker in statistics the limitations, and perhaps one may say, the faults of Mr. Keynes' book are more apparent.

Keynes's "psychological" notion of probability as a measure of "the 'degree of rational belief' to which a proposition is entitled in the light of given evidence" was useless. "To the statistician probability appears simply as the ratio which a part bears to the whole of a (usually infinite) population of possibilities."

Fisher avoided this psychological notion in his theory and practice and indeed he (Fisher, 1921, p. 24) argued against Bayes:

> Bayes (1763) attempted to find, by observing a sample, the actual probability that the population value lay in any given range. ... Such a problem is indeterminate without knowing the statistical mechanism under which different values of r come into existence; it cannot be solved from the data supplied by a sample, or any number of samples, of the population.

In 1922 Fisher considered a version of the Wrinch–Jeffreys problem. Against Bayes "postulate", i.e. that it is reasonable to assume that the *a priori* distribution of the parameter p, the probability of a success in the binomial distribution, is uniform, he (pp. 324–325) argues that, "apart from evolving a vitally important piece of knowledge, that of the exact form of the distribution of p, out of an assumption of complete ignorance, it is not even a unique solution." A uniform prior for θ, defined as $\sin\theta = 2p - 1$, leads to a posterior distribution inconsistent with that obtained from specifying a uniform prior for p. This is a variant of the von Kries paradox discussed by Keynes; see §1.1 above.

Jeffreys wrote to Fisher for a copy of this paper—the only evidence of direct communication between them before 1933. Jeffreys never referred to the "indeterminacy" argument but he reported on the paper's account of probability in *Scientific Inference* (p. 220n): "R.A. Fisher [(1922)] with what looks like the courage of despair, says that in a 'hypothetical infinite population' the ratio is perfectly definite." Fisher considered the foundations of *his* probability in good order and only identified probability as an issue in the 1950s.

Fisher recognised a determinacy to match the "indeterminacy": if there is "a super-population of known specification", the inverse argument produces a "perfectly definite value" for the probability distribution of the parameters (1930, pp. 530–531). With the introduction of the fiducial argument in his "Inverse Probability" (1930), it became important to Fisher to clarify the relationship between these two valid ways of attaching probabilities to parameters. See Aldrich (2000) for this first 1930 attempt. In the next section we see how Fisher continued to explore this relationship in his paper criticising Jeffreys's "On the Theory of Errors and Least Squares".

2.2 The Controversy

The controversy has been analysed by Lane (1980) and Howie (2002) from the point of view of the protagonists' interpretation of probability. The present interest is more in the way the controversy affected the course of Jeffreys's research. The controversy was set off when J.B.S. Haldane published a paper (1932) on the sampling problem. He did not mention Jeffreys's work on the same problem but he more than mentioned Fisher, interpreting maximum likelihood theory as inverse theory, a delicious circularity as Fisher (1922) had been correcting Pearson's confused Bayesian reasoning from the 1890s. The exasperated Fisher (1932, p. 260) wrote, "I had hoped that it should be clear that my work was based not on the tacit assumption of equal *a priori* probability but upon the explicit rejection of this assumption." However he (p. 258) had one new point, based on Haldane's observation that by taking a large enough sample the influence of the prior on the posterior can be made arbitrarily small:

> The obvious objection . . . is that if the [prior] is in reality irrelevant to our conclusions, it should have no place in our reasoning; and that if the form of our reasoning requires its introduction, the fault lies with our adoption of this form of reasoning.

This struck against the main finding of "Aspects" on the problem of sampling induction.

Jeffreys (1933a) also responded negatively to Haldane, mainly to an "absurdity" Haldane found when using a uniform prior. But then in winding up he (p. 86) replied to Fisher with a "fundamental objection":

> The only sense in which $f(r)$ is irrelevant is that when the sample is large enough the influence of the variation of $f(r)$ on the distribution of posterior probability can be made arbitrarily small; in practice statisticians take large samples for another reason, namely to reduce the probable error of r/n and thereby also minimize the displacement of the most probable value due to the variation of $f(r)$. But the fact that there are two reasons for a given procedure is no reason for inferring that one of them is meaningless or invalid. The whole reason for attaching any importance to Fisher's "likelihood' is that it is proportional to the posterior probability given by Laplace's theory, and it has no meaning outside the original sample except in terms of that theory.

Jeffreys seems to have misunderstood what Fisher the "statistician" was saying but the last sentence brings out the extent of their disagreement.

Meanwhile Fisher (1933) was criticising Jeffreys's least squares paper (§1.3 above). Fisher focussed on the prior for h and, by ignoring the t work, put back Jeffreys's education. Jeffreys seems to have first learnt of 'Student's' work only in 1937—from Wishart. (see §3.3 below).

Fisher (1933, p. 343) re-posed the question about the prior required to generate an appropriate posterior as a question about a super-population:

> What distribution *a priori* should be assumed for the value of h, regarding it as a variate varying from population to population of the ensemble of populations which might have been sampled?

This question was not one that Jeffreys had asked but perhaps Fisher thought it was the only meaningful question that could be asked. He could certainly have a bit of fun with it.

In 1922 Fisher had registered the futility of the exercise of evolving "a vitally important piece of knowledge ... out of an assumption of complete ignorance". Now he (pp. 343–344) noted "That there should be a method of evolving such a piece of information by mathematical reasoning only ... would be in all respects remarkable ... "; further, the proof can scarcely establish all that is claimed "since there is nothing to prevent our setting up an artificially constructed series of populations having any chosen distribution of h ... in which case Jeffreys's proof would certainly lead to a false conclusion."

Fisher did not stop at these general considerations but tried to expose the "fallacy" in Jeffreys's argument. Jeffreys had argued from a "fact otherwise known about the posterior probability" to the form of prior. Fisher saw no reason to accept the fact: "All that we really know ... is that *on the average of all values of* v [twice the difference between the first and second observation] the probability is exactly one-third." Thus his (p. 344) account of the steps of Jeffreys's argument (quoted in §1.3 above) runs on:

> (d) the expression so obtained is equated to 1/3, *without averaging it for all possible pairs of initial observations*; had this essential step been taken, the equation would have degenerated to an identity for all possible distributions *a priori*.

Fisher's discussion is confusing because he (p. 344) headlines the fallacy as "assuming that the probability shall be $1/3$, *independently of the distance apart of the first two observations*" but Jeffreys did not assume this, nor did Fisher's exposé involve it.

The second half of Fisher's paper did not directly address anything Jeffreys had written but continued the discussion in "Inverse probability" (1930) on the difference between super-population reasoning and the fiducial argument, transposing the fiducial argument from the correlation coefficient to σ (replacing h for the "convenience" of statisticians). Fisher (1933, p. 347) starts from the observation that, as the ratio s/σ is independent of unknown parameters, "we can assert, without reference to any unknown quantities, or to their unknown probabilities *a priori*, with what frequency any particular value of the ratio s/σ will be exceeded in random samples." He proceeds to invert statements about s so they become statements about σ. Thus a statement of the form

$$s > s_{0.01}(\sigma) \tag{5}$$

> is equivalent to the inequality,

$$\sigma < \sigma_{0.99}(s) \tag{6}$$

> since for any value of the probability chosen, the corresponding values of s and σ increase together from 0 to ∞.
> Now we know that inequality (5) will be satisfied in just 1 per cent. of random trials, whence we may infer that the inequality (6) will be satisfied with the same frequency.

Fisher insists that the statement about (6) is *not* a statement of inverse probability.

> This distinction is necessary since the assumption of a given frequency distribution *a priori* ... might conceivably be true .... The probabilities differ in referring to different populations; that of the fiducial probability is the population of all possible random samples, that of the inverse probability is a group of samples selected to resemble that actually observed.
> It is the lack of this distinction that gives a deceptive plausibility to the frequency distribution *a priori*

$$df = d\sigma/\sigma = d(\log \sigma).$$

> For this particular distribution *a priori* makes the statements of inverse and fiducial probability numerically the same, and so allows their logical distinctness to be slurred over.

The argument could be turned round: the fiducial argument can be justified as an inverse argument with *this* prior and Jeffreys would make such a case in the *Probability*—see §3.3 below. However, Fisher (p. 348) argued that this prior could not support anything.

> This particular distribution *a priori* is, however, not only hypothetical but unacceptable as such, since it implies that all ranges of values of the parameter covering finite ratios, however great, are infinitely improbable.

Fisher also drew attention to Jeffreys's timidity in pressing his argument:

> Jeffreys himself seems to feel some doubt as to the general validity of the distribution ... for he says: "The solution must break down for very small h ... and for large h ..."; though he does not indicate in what way his mathematical proof fails for these parts of the range. (p. 344)

In his reply Jeffreys (1933b) returned to the key points of *Scientific Inference* because he could "find no basis of agreement to use as the starting-point in a reply to his criticisms". The paper is half-over before he reaches Fisher. When he does, he defends his analysis against Fisher's criticism and then criticises Fisher's own work.

When pressed Jeffreys actually became less tentative. *Scientific Inference* had not fully endorsed the $1/h$ prior but (p. 67) proposed a modification to make its integral unity to conform to the "convention" (p.12) that "1 is the constant to be attached to certainty." Now, however, impropriety became a virtue. Jeffreys (1933b, p. 531) gave two new arguments to support "*the fact ... that my distribution is the only distribution of prior probability that is consistent with complete ignorance of the value of h.*". The ratio of the probabilities that h is less than and greater than h_0 is indeterminate when the distribution is $1/h$; this is required if the distribution "tells us nothing as to the probability that h will exceed any given value". The second argument (p. 352) is that "a single observation does not affect the distribution of the probability of h. This is what we should expect because a single observation can tell us nothing about its own precision."

Jeffreys's counter-attack (p. 532) contains a general point about sample-space averaging which became an important part of his case against "direct methods" (§3.3 below):

> Fisher proceeds to reduce my theory to absurdity by integrating with respect to all values of the observed measures. This procedure involves a fundamental confusion, which pervades the whole of his statistical work and deprives it of all meaning. The essential distinction in the problem of inference is the distinction between what we know and what we are trying to find out: between the data and the proposition on the data we are trying to find out. If we have made two observations, $\pm a$ in my notation ... those are our observations and there is nothing more to be said. To integrate with respect to them and average a function of them over the range of integration is an absolutely meaningless process. Yet in Fisher's constructive, as well as in his destructive work, this process is carried out again and again.

Jeffreys commented on a sample of Fisher's "constructive" work, the fiducial argument, that it "refers to a mean taken over all possible random samples, and it may be asked why this should be thought to have much relevance to any particular sample." (p. 534). It is striking that Jeffreys's first discussion of sample space averaging was occasioned by this exotic application and not by the arguments he would have found mixed with Bayesian arguments in "the usual presentation of the theory of errors".

Jeffreys and Fisher had simultaneous last words for the journal had them coordinate their responses. The final papers describe the author's general position and comment on the controversy and how the other had misinterpreted him. Fisher's "Probability likelihood and quantity of information in the logic of uncertain inference" accused Jeffreys of making the first of these do the job of all three. Jeffreys would have his say on this matter in the *Probability*—see §3.1 below—but perhaps the most telling sentence in his reply was, "I find it difficult to answer criticisms of the theory, because most of them seem to refer to something different from what I intend, and I cannot see what." (1934, p. 10).

Lane (p. 159) sums up, "Nothing was settled by the exchange. Neither scientist seems to have convinced the other of anything." They certainly left with their convictions reinforced and their sense of authority intact. Jeffreys went away sharper for the only criticism, that moved him, moved him to a more extreme position. Yet the exchange had its educational side. Jeffreys was forced to go beyond the unacceptable premisses and examine frequentist reasoning and he was introduced to a body of work that went significantly beyond the physicists' "combination of observations". The project of using the theory of probability to produce better "combinations" could be subsumed in the grander scheme of re-founding statistics.

2.3 "The Best Justification"

In December 1934 Fisher (1935b) read "The logic of inductive inference"—an account of "some of the theoretical researches with which I have been associated."—to the RSS. Jeffreys made a written contribution to the discussion. Predictably they disagreed about probability. Fisher was developing his view that Bayes framed the valid super-population argument and deserved credit for *not* publishing the invalid argument, unlike the rash Laplace who lacked "Bayes' scientific caution" (Fisher, 1936, p. 247). Jeffreys (1935a, p. 70) enquired, "I should be interested to know the source of Professor Fisher's remark that in the theory of inverse probability the method was to introduce a postulate concerning the population from which the unknown population was supposed to be drawn."

Jeffreys (1935a, p. 72) still objected to the fiducial argument:

> it shows insufficient respect to the observed data and does not answer the right question. When the actual difference in the two sampling ratios is given exactly, the possibility that a greater difference might have been obtained seems irrelevant; but actually it is these greater differences that contribute most of the fiducial probability. I think that when a question is proposed . . . the questioner . . . wants the answer in terms of the posterior probability with respect to the observed data. If he accepts the fiducial probability as an answer it is because he mistakenly interprets it as a posterior probability.

However Jeffreys gave a more positive shading to some of his old criticisms: Fisher "seems to set up his use of likelihood in opposition to the theory of probability" but the theory "provides the use of likelihood with its best justification"; furthermore Fisher's argument about information and sufficiency "would be made much easier by an explicit use of probability".

Already in February 1934—see (Bennett, 1990, p. 150)—Jeffreys had told Fisher, "My quarrel is not with you, but with Eddington and similar people." In a piece in *Nature* (1937b, p. 1004) Jeffreys criticised the *a priori* view of physical laws advanced by some physicists. His own system of induction and the systems given by Fisher and "other statisticians" are, despite their differences, satisfactory. Fisher might equally have said "My quarrel is not with you, but with Karl Pearson". Jeffreys managed to charm his way into contributing a paper to Fisher's *Annals of Eugenics*. In July 1937 Jeffreys mentioned Pearson's "stickiness" towards maximum likelihood (see Bennett (1990,

p. 164)), though the published paper (1938, p. 146) begins, "Prof. R.A. Fisher ... does not point out what appears to me the fundamental inconsistency in Pearson's position", i.e. using the method of moments while accepting the inverse argument. Jeffreys also thought (Bennett, 1990, p. 164) the paper could correct the idea that "seems to have grown up" that maximum likelihood and inverse probability are "opposed"—an outrageously disingenuous proposal, given it was the editor's idea. In his note Fisher (1938, p. 151) praises Jeffreys as one who accepts the "traditional doctrine" yet who has "constantly endeavoured to bring it into relation with the problems of practical research", not to speak of his openness to the "newer ideas of mathematical statistics".

3 The *Probability* and "Modern Statistics"

Jeffreys was now writing statistical papers (e.g. 1935b, 1936, 1937a). In 1937 came an enlarged *Scientific Inference* (1931a) and in 1939 came the *Probability*. This very ambitious work developed the probability part of *Scientific Inference* to encompass theory, methods and alternative standpoints. It re-wrote modern statistics in terms of "the theory of probability", the theory of inductive inference founded on the principle of inverse probability. Applications to physics were prominent but the argument was followed wherever it led, e.g. to randomization in experiments. Lindley (1986) has given a modern subjectivist's "re-reading".

I will discuss its principles of "estimation" and "significance tests" and its account of the fiducial argument. The fiducial argument and the use of tail areas in testing are items two and three in Jeffreys's (1939, p. 323) list of his theoretical disagreements with Fisher:

> My main disagreement ... concerns the hypothetical infinite population ... Another is that, as in the fiducial argument, an inadequate notation enables him, like 'Student', to pass over a number of really difficult steps without stating what hypotheses are involved in them. The third is the use of the P integral, but Fisher's alertness for possible dangers is so great that he has anticipated all the chief ones.

Jeffreys (1939, p. 324) disapproved of Fisher's way of meeting these dangers by generating "independent postulates" but he admired Fisher's practice. The passage continues

> I have in fact been struck repeatedly in my own work, after being led on general principles to a solution of a problem, to find that Fisher has already grasped the essentials by some brilliant piece of common sense, and that his results would be either identical with mine or would differ only in cases where we should both be very doubtful.

Jeffreys had remarked upon the combination of practical agreement and theoretical disagreement already in a letter to Fisher of May 1937: "only once in a blue moon" would we disagree about the "inference to be drawn in any particular case" (Bennett, 1990, p. 162).

3.1 "Problems of Estimation"

Jeffreys divides the problems of statistical inference into two. A problem of estimation (p. 94) is "one where we are given the form of the law, in which certain parameters can be treated as unknown, no special consideration needing to be given to any particular values, and we want the probability distribution of these parameters given the observations." In significance testing *special consideration* is given to a particular value: "our problem is to compare a suggested value of a new parameter, often 0, with the aggregate of other possible values" (p. 193).

Chapter III, the main chapter on estimation lays down rules for choosing a prior and then derives posterior distributions for the parameters of several laws familiar to statisticians. It is essentially an extension of the treatment of the sampling problem and the theory of errors in *Scientific Inference*. The chapter begins (pp. 96–103) by considering how we should say "that the magnitude of a parameter is unknown, when none of the possible values need special attention". Two rules are proposed: that the prior for the parameter should be taken as uniform or, if the parameter is positive, its logarithm should be taken as uniformly distributed.

These rules require an alteration in the foundations—the waiving of a "convention". Chapter I on *Fundamental notions* develops probability on the basis of six axioms and three conventions. The waiver affects convention 3 (p. 21)

> CONVENTION 3. If p entails q, then $P(q \mid p) = 1$.
> This is the rule generally adopted; but there are cases where we wish to express ignorance over an infinite range of values of a quantity, and it is then convenient to express certainty that the quantity lies in that range by ∞, in order to keep ratios for finite ranges determinate. None of our axioms so far has stated that we must always express certainty by the same number on different data, merely that we must on the same data; but with this exception it is convenient to do so.

The waiver formalised a change he made in the course of the controversy with Fisher.

In "modern statistics", according to Fisher (1922 and 1925), estimation is concerned with the efficient extraction of information about a parameter from a sample. Jeffreys covers Fisherian extraction with his own version of concepts like sufficiency—see Aldrich (2002, §3)—but he makes no great fuss about it. In "modern statistics", according to Neyman (1935, p. 75), estimation is based on the "frequency of errors in judgement"; this provides a "sufficiently simple and unquestionable principle in statistical work" while the "amount of information is too complicated and remote to serve as a principle". Jeffreys does not say anything systematic about what is to be done with the "probability distribution" when it is found. It will support intervals statements but there is no general discussion of what should be reported as *the* point estimate. For the Poisson distribution he (p. 115) gives the expectation of the parameter given the data and notes that the "maximum probability density is at a slightly smaller value." In the important least squares case (p. 129) the mean and mode of the posterior density are equal and equal to the least squares value.

The subsidiary chapter on "approximate methods and simplifications" presents maximum likelihood as an approximation to a Bayesian maximum posterior concluding that "the errors introduced by treating the prior as uniform will be of no practical importance if the number of observations is large" a thought going back to Wrinch & Jeffreys (1917). Curiously Jeffreys (1939, pp. 146ff) presents minimum χ^2 as "an approximation to maximum likelihood" when there are no sufficient statistics and maximum likelihood is "very laborious", although the ultimate payoff is its approximation to the Bayesian solution.

3.2 Significance Testing and the P Integral

The foundations for Jeffreys's theory of significance testing were laid in "Some tests of significance treated by the theory of probability" (1935b) and "Further tests of significance" (1936). The first begins from a doubt about "the usual procedure" of judging a difference significant if "it exceeds a certain rather arbitrary multiple of the standard error" and its aim is to bring testing within the scope of the "theory of probability". The usual procedure was an improvisation based on the use of the posterior distribution appropriate to estimation and did not reflect the belief structure of the tester. The novelty in Jeffreys's approach was in the specification of a prior for testing.

The 1935–6 papers described the purpose of testing and worked through examples from the physical science and then from the statistics literature. This material—with many additions—became Chapters V and VI of the *Probability*. For Jeffreys (1939, p. 95) the function of significance tests is "to provide a way of arriving, in suitable cases, at a decision that at least one new parameter is needed to give an adequate representation of the data and valid inferences to future ones." Such tests operated against a background of laws ordered by degree of simplicity (p. 194). Jeffreys's treatment of significance testing was integrated with his view of science as a whole.

The preliminary sketch (chapter V, pp. 193–195) indicates what is involved:

> Our problem is to compare a suggested value of a new parameter, often 0, with the aggregate of other possible values. We do this by stating a hypothesis q, that the parameter has the suggested value, and $\backsim q$, that it has some other value to be determined from the observations. q would always be what Fisher calls the *null hypothesis* . . . . The essential feature is that we express ignorance of whether the new parameter is needed by taking half the prior probability for it as concentrated on the value indicated by the null hypothesis, and distributing the other half uniformly over the range possible.

The inferentially relevant quantity is the posterior odds ratio, viz.

$$K = \frac{P(q\,|\theta H)}{P(\backsim q\,|\theta H)}$$

where θ is the "observational evidence" and H is a statement of background knowledge (p. 47). Jeffreys's treatment is based on even prior chances for q and $\backsim q$ but he also envisages situations where these are not even and then K corresponds to the Bayes factor of the modern literature. Appendix I has tables of K for the important χ^2 and t cases.

The 1935–6 papers had not mentioned Fisher (or Neyman and Pearson) and the book's critique of testing using direct methods was largely new. It begins in the preface (p. v). For the most part modern statisticians

> have rejected the notion of the probability of a hypothesis, and thereby deprived themselves of any way of saying precisely what they mean when they decide between hypotheses. . . . [M]ost statisticians appear to regard observations as a basis for possibly rejecting hypotheses, but in no case for supporting them . . . . [This] attitude if adopted consistently would reduce all inductive inference to guesswork.

The allusion is to Fisher's statement in *The Design of Experiments* (1935a, p. 19), "it should be noted that the null hypothesis is never proved or established, but is possibly disproved in the course of experimentation."

Unlike Fisher, Jeffreys required a "statement of the alternative" as well as of the null. Not surprisingly he (p. 326) praised this feature of the Neyman–Pearson (1933) approach, although he thought that in the case of simple hypotheses a straight consideration of the likelihood ratio would be preferred to a tail area and, where there is an adjustable parameter, "the total risk of errors of the second kind must be compounded of the power functions over the possible values with regard to their risk of occurrence." Nevertheless he (pp. 326–327) found it "interesting" to consider the relationship between his method and that of Neyman and Pearson by finding the value of K that minimises the total number of mistakes of both kinds on the assumption that "world-frequencies" are in proportion to the prior probability used to express ignorance. The value of K is unity.

Jeffreys found it hard to satisfy himself on how the Student problem should be treated. The *Probability* (1939, p. 198) starts out, "The simplest case when an unknown standard error has to be found from the data is that of a set of measures that, on q, would be derived from a normal law centred on 0, but on $\backsim q$ will refer to one centred on a value α different from 0." Jeffreys actually made three attempts: in the 1935b paper and in the 1st and 2nd editions of the *Probability.* In the 1st edition he took as his "fundamental" unknowns α, the mean, and the "whole variation from 0", i.e.

$\alpha^2 + \sigma^2$, which on his principles led to the priors

$$P(q d\sigma \mid H) \propto d\sigma/\sigma$$
$$P(\backsim q\ d\sigma d\alpha \mid H) \propto d\sigma d\alpha/2\sigma^2.$$

After some approximating Jeffreys (p. 201) obtains

$$K = \left(\frac{2n}{\pi}\right)^{\frac{1}{2}} \left(1 + \frac{a^2}{s'^2}\right)^{-\frac{(n-3)}{2}}$$

where a is the sample mean and s'^2 the sample variance.

Jeffreys (p. 202) established that the posterior behaviour for $n = 1$ was appropriate: "for a single observation $K = 1$, and the test is indeterminate. This what we should expect since we have then no means of separating systematic and random variations whatever the magnitude of the measure may be." Subsequently he (1942, 1948) considered a further set of facts on the posterior, including the requirement that when $n \geq 2$, $s' = 0$ and $\overline{x} \neq 0$ should imply $K = 0$ (1948, p. 243). The simplest expression of his prior beliefs and facts about the posterior involved the following conditional prior for x:

$$P(dx \mid \backsim q\ \sigma H) \propto \frac{1}{\pi(1 + x^2/\sigma^2)} \frac{dx}{\sigma},$$

writing the expression of (1948, p. 243) in the notation of 1939.

At the beginning of his critique of "current statistical theory" Jeffreys (1939, p. 300) insists, "Their [statisticians'] practice, when they come to specific applications, is mostly very good; the fault is in the precepts." The leading instance was the use of the "P integral" in significance testing. (P was the conventional symbol for the tail area, the modern "prob-value".) The use of the P integral was an example of averaging over the sample space which, from his first encounter with Fisher (§2.2 above), Jeffreys had judged to be "an absolutely meaningless process". In the *Probability* he (p. 315) states that "it is with regard to this use of P that I differ from all the present statistical schools and detailed attention to what it means is needed." Jeffreys considered his approach and that of the statisticians to be based on the same fundamental idea: "that a law should not be accepted on data that themselves show large departures from its predictions." However their implementation of the idea (p. 315) made no sense to him

> If P is small, that means that there have been unexpectedly large departures from prediction. But why should these be stated in terms of P? The latter gives the probability of departures, measured in a particular way, equal to *or greater than* the observed set, and the contribution from the actual value is nearly always negligible. *What the use of P implies, therefore, is that a hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred.* This seems a remarkable procedure. On the face of it the fact that such results have not occurred might more reasonably be taken as evidence for the law, not against it. The same applies to all the current significance tests based on P integrals.

In Appendix I where K values for some standard situations are presented Jeffreys compares the use of K with that of significance levels. "Users of these tests [based on the P integral] speak of the 5 per cent. point in much the same way as I should speak of the $K = 10^{-\frac{1}{2}}$ point." (p. 360). At moderate sample sizes the points are "not very different" but he recognises that "At large numbers of observations there is a difference, since the test based on the integral would sometimes assert significance at departures which would actually give $K > 1$." This possibility of discrepancy was later headlined by Lindley (1957) as "A Statistical Paradox".

3.3 The Fiducial Argument

While Jeffreys consistently criticised the reasoning behind the P integral, he changed his mind about the fiducial argument. Initially (see §2.2 and §2.3 above) he thought it was fallacious, a natural reaction given Fisher's (1930, 1933) way of presenting it as a by-product of (P integral) significance testing. However, Jeffreys came to the conclusion that "In fact the fiducial argument, when completed, and the inverse argument, are simply different ways of saying the same thing; the hypotheses are identical and so are the results." (1940, p. 49). If he had a tutor here, it was 'Student'—W.S. Gosset.

In *Scientific Inference* (see §1.3) Jeffreys derived, using inverse probability, a form identical to the one 'Student' (1908) had derived by direct methods. Surprisingly 'Student's' work had *not* come up in the controversy with Fisher—indeed Jeffreys (1939, p. 323) recalled, "I thought he [Fisher] was attacking the 'Student' rule . . . ". In 1937 Jeffreys wrote, "it appeared to me that the identity of the results in form must be accidental. It turns out, however, that there is a definite reason why they should be identical, and that this throws a light on the use of direct methods for estimation and on their relation to the theory of probability." (1937a, p. 326.)

The 1931a result was the following expression for the posterior distribution of the mean x given the sample mean, $\overline{x}$, and standard deviation, σ':

$$P(dx\,|\overline{x}, \sigma', k) \propto \left\{1 + \frac{(x - \overline{x})^2}{\sigma'^2}\right\}^{-\frac{1}{2}n} dx$$

where "previous knowledge" k expresses "the truth of the normal law but nothing about x and σ." (Jeffreys used the same $\frac{1}{n}$ form for the sample standard deviation as 'Student' and he refers to 'Student's' z rather than to the t form introduced by Fisher). A change of variable to z, defined by $x - \overline{x} = z\sigma'$, leads to

$$P(dz\,|k) \propto \left\{1 + z^2\right\}^{-\frac{1}{2}n} dz$$

where $\overline{x}$ and σ' have been suppressed on the left side because they do not appear on the right.

Jeffreys then gives a direct argument beginning from the joint density of $\overline{x}$ and σ' and leading to the same expression for dz. Jeffreys wants to understand how 'Student's' z distribution obtained in this way could be used as a posterior distribution: what hidden assumptions would make this possible? It was detective work like that involved in identifying $\frac{1}{h}$ as the implicit prior in another non-standard probability assessment—discussed above in §1.3. It may seem a pointless investigation given that the z distribution was used mainly in significance testing. Jeffreys thought that, whatever they might say, statisticians *must* be making probability statements but he could also point to a probability statement about the mean in 'Student' (1908, p. 13): "if two observations have been made and we have no other information, it is an even chance that the mean of the (normal) population will lie between them." Moreover Fisher had been using the t distribution in the fiducial argument for the normal mean since his (1935c). Fisher (1939, pp. 4–5) in his obituary of 'Student' interpreted the 1908 passage as hinting at the fiducial development.

The substance of the 1937 analysis went into the *Probability* (1939, pp. 309–313). Jeffreys (p. 310) summarised. and commented

> But we must notice that it involves two hypotheses: first, that nothing in the observations but x and s [replacing σ'] is relevant; secondly, that whatever they may be in the actual observations we are at full liberty to displace or rescale the distribution to any extent. The first is perhaps natural, but it is desirable to keep the number of hypotheses as small as possible, whether they are natural or not, and the result is proved by the principle of inverse probability. The second can mean only one thing, that the true value x and the standard error σ are completely unknown.

Jeffreys (p. 311) reports (see also Bennett (1990, p. 270)) that "'Student' called my attention to the vital word ["unique"] just after the publication [of Jeffreys (1937a)] showing that he had in fact clearly noticed the necessity of the condition that the sample considered must constitute our only information about x and σ." There was indeed a note in 'Student' (1917, p. 414): "By unique I mean to say that all the information we have (or at all events intend to use) about the distribution of the population is given by the sample in question." In his 1908 paper 'Student' moved easily from sampling distributions to the probability of hypotheses: "From the table the probability is 0.9985, or the odds are about 666 to 1 that 2 is the better soporific." (p. 21). He was an unorthodox Bayesian for elsewhere he put the sampling distribution of a statistic in the place of the likelihood. See Aldrich (1997, p. 166.)

Jeffreys (p. 311) states that "similar considerations affect Fisher's fiducial argument", noticing how "in speaking of the probability distribution of μ in the light of the sample Fisher has apparently abandoned the restriction of the meaning of probability to direct probabilities." Now that he and Fisher are giving the same solution to the same problem but following different routes, Jeffreys (p. 312) could say

> My only criticism of both his argument and 'Student's' is that they omit important steps, which need considerable elaboration, and that when these are given the arguments are much longer than those got by introducing the prior probability to express previous ignorance at the start.

The *Probability* was not the end of Jeffreys's involvement with the fiducial argument. He discussed the two-sample Behrens–Fisher problem in correspondence with Fisher and, then, in a paper (1940) published in Fisher's journal. Fisher, feeling very isolated, wrote a paper (1941), that was both learned and angry, insisting that his treatment of the Behrens–Fisher problem was an extension of that of 'Student's' problem. Fisher (1941, p. 142) mentioned Jeffreys, "whose logical standpoint is very different from my own", and appealed to his work for support: "Jeffreys (1940) and Yates (1939), from entirely different standpoints, have explained the logic of the argument, and its close analogy, or more properly identity, with that required for 'Student's' original test." (p. 149). Stone (1983) and Zabell (1992) consider the rather radical reconstruction of the fiducial argument that Fisher effected around this time.

4 The *Probability* and the Statisticians

Writing the *Theory of Probability* and the dozen or so papers around it did not make Jeffreys a statistician, or at least did not gain him recognition as one. Wilks (1941, p. 194) was prophetic in thinking it "doubtful that there will be many scholars thoroughly familiar with the system of statistical thought initiated by R.A. Fisher and extended by J. Neyman, E.S. Pearson, A. Wald and others who will abandon this system in favour of the one proposed by Jeffreys in which inverse probability plays the central role." Statisticians may not have entirely agreed with Fisher's saying, "[Jeffreys] makes a logical mistake on the first page which invalidates all the 395 formulae in his book" (from Box (1978, p. 441)) but they acted as if they did. Jeffreys studied "direct" arguments, either to replace them or to accept them as approximations to the inverse solution, but statisticians felt no obligation to study his arguments. They already knew how to proceed and anything that was new to them was based on premisses they could not accept.

In 1939 there were still Bayesian survivors of the pre-Fisher era, like Bowley, but they did not look to Jeffreys for support, nor did he recognise them. In the event he wrote the *Probability* to please himself. Because statisticians did not engage with its arguments the only impulse to change the book came from himself. Jeffreys continued to refine the analysis after publication (see §3.2 above) and in 1946 he introduced his "invariant form for the prior probability in estimation problems" which filled a gap first identified in 1931 (see §1.3 above). The substantial additions in the 2nd (1948) and 3rd (1960) editions of the *Probability* were not in response to criticism from statisticians but in response to Jeffreys's dissatisfaction with his own earlier efforts.

If Jeffreys was writing for anybody else, it was for Fisher. But, while Fisher was the central figure in statistics in 1933, the centre was moving. The Neyman, Pearson and Wald "extensions" of Fisher were defining a new centre and Jeffreys, in addressing Fisher, was missing it. The Bayesian revival, when it came, came from extensions of the extensions, as in Raiffa & Schlaifer (1961), or from immanent criticism of the extensions, as in Savage (1954). Already in 1939 the *Probability* resembled a dynosaur. Its probability was pre-Kolmogorov and its logical point of reference was *Principia Mathematica* which was written before the Great War. For post-war statisticians Cramér's *Mathematical Methods* (1946) was a more powerful model of how to apply mathematics. The book's integrity and architectonic grandeur counted against it. To follow Jeffreys's treatment of significance tests it was necesary to go back to the simplicity postulate and the issues of scientific method associated with it.

Finally, what of Fisher? When writing about Jeffreys, Fisher could be teasing or respectful. In his review of the second edition of *Scientific Inference* he (1957, p. 595) wrote, "Its basis, a series of papers by the author and the brilliant Miss D. Wrinch, excited interest and some respect for its daring, but so far as the reviewer knows, won no adherents." Yet anyone recognising "the rational cogency of the fiducial form of argument." (Fisher, 1956, p. 56) was entitled to respect. Fisher never admitted learning anything from Jeffreys but commentators, including Lane (1980, p. 159) and Zabell (1992, pp. 378 & 386), have remarked on how Fisher grew to resemble Jeffreys, in particular, how his later formulations of the fiducial argument incorporated points Jeffreys had made against him.

Acknowledgement

I am grateful to the Editor and the referees for their comments.

References

Aldrich, J. (1997). R.A. Fisher and the making of maximum likelihood 1912–1922. *Statist. Sci.*, **12**, 162–176.
Aldrich, J. (2000). Fisher's "Inverse Probability" of 1930. *International Statistical Review*, **68**, 155–172.
Aldrich, J. (2002). How likelihood and identification went Bayesian. *International Statistical Review*, **70**, 79–98.
Aldrich, J. (2003a). *Harold Jeffreys as a Statistician*. Website
http://www.economics.soton.ac.uk/staff/aldrich/jeffreysweb.htm
Aldrich, J. (2003b). *A Guide to R.A. Fisher*. Website
http://www.economics.soton.ac.uk/staff/aldrich/fisherguide/rafreader.htm
Bennett, J.H. (1990). (Ed.) *Statistical Inference and Analysis: Selected Correspondence of R.A. Fisher*. Oxford: Oxford University Press.
Box, J.F. (1978). *R.A. Fisher: The Life of a Scientist*. New York: Wiley.
Broad, C.D. (1918/20). The relation between induction and probability I–II. *Mind*, **27**, 389–404 & **29**, 11–45.
Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press: London.
Fisher, R.A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, **1**, 3–32.
Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society, A*, **222**, 309–368.
Fisher, R.A. (1922/3). Review of J.M. Keynes's Treatise on Probability. *Eugenics Review*, **14**, 46–50.
Fisher, R.A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.
Fisher, R.A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society*, **26**, 528–535.
Fisher, R.A. (1932). Inverse probability and the use of likelihood. *Proceedings of the Cambridge Philosophical Society*, **28**, 257–261.
Fisher, R.A. (1933). The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proc. Roy. Soc. (London) A*, **139**, 343–348.
Fisher, R.A. (1934). Probability, likelihood and the quantity of information in the logic of uncertain inference. *Proc. Roy. Soc. (London) A*, **146**, 1–8.
Fisher, R.A. (1935a). *The Design of Experiments*. Edinburgh: Oliver & Boyd.
Fisher, R.A. (1935b). The logic of inductive inference (with discussion). *J. Roy. Statist. Soc.*, ***98**, 39–82.*
Fisher, R.A. (1935c). The fiducial argument in statistical inference. *Annals of Eugenics*, **6**, 391–398.

Fisher, R.A. (1936). Uncertain inference. *Proceedings of the American Academy of Arts and Science*, **71**, 245–258.

Fisher, R.A. (1938). Comment on H. Jeffreys' paper on maximum likelihood, inverse probability and the method of moments. *Annals of Eugenics*, **8**, 151.

Fisher, R.A. (1939). 'Student'. *Annals of Eugenics*, **9**, 1–9.

Fisher, R.A. (1941). The asymptotic approach to Behrens's integral, with further tables for the d test of significance. *Annals of Eugenics*, **11**, 141–172.

Fisher, R.A. (1956). *Statistical Methods and Scientific Inference*. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1957). Review of *Scientific Inference* by Sir Harold Jeffreys. *Cambridge Review*, 18th May, 595–597.

Geisser, S. (1980). The contributions of Sir Harold Jeffreys to Bayesian inference, pp. 13–20 of Zellner (Ed.) (1980).

Good, I.J. (1980). The contributions of Jeffreys to Bayesian statistics, pp. 21–34 of Zellner (Ed.) (1980).

Haldane, J.B.S. (1932). A note on inverse probability. *Proceedings of the Cambridge Philosophical Society*, **28**, 55–61.

Howie, D. (2002). *Interpreting Probability: Controversies and Developments in the Early Twentieth Century*. New York: Cambridge University Press.

Irwin, J.O. (1941). *Theory of Probability* by Harold Jeffreys. *J. Roy. Statist. Soc.*, **104**, 59–64.

Jeffreys, H. (1916). On certain possible distributions of meteoric bodies on the solar system. *Monthly Notices of the Royal Astronomical Society,* **77**, 84–112.

Jeffreys, H. (1922). The theory of probability [review of Keynes (1921)]. *Nature*, **109**, 132–133.

Jeffreys, H. (1924). *The Earth*. (2nd edition 1929) Cambridge: Cambridge University Press.

Jeffreys, H. (1931a). *Scientific Inferenc*e. Reprinted with additions 1937 and followed by a 2nd. edition 1957. Cambridge: Cambridge University Press.

Jeffreys, H. (1931b). The revision of seismological tables. *Monthly Notices of the Royal Astronomical Society, Geophysical Supplement*, **2**, 329–348.

Jeffreys, H. (1932a). An alternative to the rejection of observations. *Proc. Roy. Soc. (London) A*, **137**, 78–87.

Jeffreys, H. (1932b). On the theory of errors and least squares. *Proc. Roy. Soc. (London) A*, **138**, 48–55.

Jeffreys, H. (1933a). On the prior probability in the theory of sampling. *Proceedings of the Cambridge Philosophical Society*, **29**, 83–87.

Jeffreys, H. (1933b). Probability, statistics and the theory of errors. *Proc. Roy. Soc. (London) A*, **140**, 523–535.

Jeffreys, H. (1934). Probability and scientific method. *Proc. Roy. Soc. (London) A*, **146**, 9–16.

Jeffreys, H. (1935a). Discussion of Fisher (1935b). *J. Roy. Statist. Soc.,* **98**, 70–72.

Jeffreys, H. (1935b). Some tests of significance treated by the theory of probability. *Proceedings of the Cambridge Philosophical Society*, **31**, 203–222.

Jeffreys, H. (1936). Further significance tests. *Proceedings of the Cambridge Philosophical Society*, **32**, 416–445.

Jeffreys, H. (1937a). On the relation between direct and inverse methods in statistics. *Proc. Roy. Soc. (London) A*, **160**, 325–348.

Jeffreys, H. (1937b). Modern Aristotelianism: contribution to discussion, Supplement to *Nature*, **139**, 1004–1005.

Jeffreys, H. (1938). Maximum likelihood, inverse probability and the method of moments. *Annals of Eugenics*, **8**, 146–151.

Jeffreys, H. (1939). *Theory of Probability*. (2nd and 3rd editions in 1948 and 1960), Oxford: University Press.

Jeffreys, H. (1940). Note on the Behrens–Fisher formula. *Annals of Eugenics*, **10**, 48–51.

Jeffreys, H. (1942). On the significance tests for the introduction of new functions to represent measures. *Proc. Roy. Soc. (London) A*, **180**, 256–268.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc. Roy. Soc. (London) A*, **186**, 453–461.

Jeffreys, H. (1963). Review of *The Foundations of Statistical Inference* by L.J. Savage and others. *Technometrics*, **5**, 407–410.

Jeffreys, H. (1974). Fisher and inverse probability. *International Statistical Review*, **42**, 1–3.

Jeffreys, H. (1980). Some general points in probability theory, pp. 451–453 in Zellner (1980).

Jeffreys, H. & Bullen, K.E. (1935). Times of transmission of earthquake waves. *Bur. Centr. Seism Trav. Sci*, **11**, 3–96 & 1–106.

Johnson, W.E. (1924). *Logic, Part III. The Logical Foundations of Science,* Appendix on Eduction pp. 178–189. Cambridge, Cambridge University Press.

Johnson, W.E. (1932). Appendix (ed. by R.B. Braithwaite) to "Probability: deductive and inductive problems". *Mind*, **41**, 421–423.

Keynes, J.M. (1921). *A Treatise on Probability*. London: Macmillan. Collected Writings Edition, 1973, London, Macmillan for the Royal Economic Society.

Lane, D.A. (1980). Fisher, Jeffreys and the nature of probability. In *R.A. Fisher: An Appreciation*. Eds. S.E. Fienberg and D.V. Hinkley, pp. 148–160. New York: Springer.

Lindley, D.V. (1957). A statistical paradox. *Biometrika*, **44**, 187–192.

Lindley, D.V. (1986). On re-reading Jeffreys. In *Pacific Statistical Congress*. Eds. I.S. Francis *et al.*, pp. 35–46. New York: Elsevier.

Neyman, J. (1935). Discussion of Fisher (1935b). *J. Roy. Statist. Soc.*, **98**, 73–76.

Neyman, J. & Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London A*, **231**, 289–337.

Pearson, K. (1911). *The Grammar of Science. 3*rd ed., (1st ed. 1892). London: White.

Pfanzagl, J. & Sheynin, O. (1996). Studies in the history of probability and statistics XLIV: a forerunner of the t-distribution. *Biometrika*, **83**, 891–898.

Raiffa, H.A. & Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Boston: Graduate School of Business Administration, Harvard University.

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

Stone, M. (1983). Fiducial probability. In *Encyclopedia of Statistical Science*, **3**. Eds. S. Kotz & N.L. Johnson, pp. 81–85. New York: Wiley.
'Student' (1908). The probable error of a mean. *Biometrika*, **6**, 1–25.
'Student' (1917). Tables for estimating the probability that the mean of a unique sample of observations lies between $-\infty$ and any given distance of the mean of the population from which the sample is drawn. *Biometrika*, **17**, 414–417.
Venn, J.A. (1888). *The Logic of Chance*, 3rd ed. (1st ed. 1866). London: Macmillan.
Whitehead, A.N. & Russell, B. (1910–1913). *Principia Mathematica*. Cambridge: Cambridge University Press.
Whittaker, E. & Robinson, G. (1924). *Calculus of Observations*. Edinburgh: Blackie.
Wilks, S.S. (1941). *Theory of Probability* by Harold Jeffreys. *Biometrika*, **32**, 192–194.
Wrinch, D. & Jeffreys, H. (1919). On some aspects of the theory of probability. *Philosophical Magazine*, **38**, 715–731.
Wrinch, D. & Jeffreys, H. (1921/23). On certain fundamental principles of scientific inquiry (two papers). *Philosophical Magazine*, **42**, 369–390 & **45**, 368–374.
Wrinch, D. & Jeffreys, H. (1923). On the seismic waves from the Oppau explosion of 1921 Sept. 21st. *Monthly Notices of the Royal Astronomic Society: Geophysical Supplement*, **1**, 15–22.
Yates, F. (1939). An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters. *Proceedings of the Cambridge Philosophical Society*, **35**, 579–591.
Zabell, S. (1982). W.E. Johnson's "Sufficientness" Postulate. *Ann. Statist.*, **10**, 1091–1099.
Zabell, S. (1992). R.A. Fisher and the Fiducial Argument. *Statist. Sci.*, **7**, 369–387.
Zellner, A. (Ed.) (1980). *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*. Amsterdam: North-Holland.

Résumé

Cet article a pour but d'examiner les travaux en statistique du physicien Harold Jeffreys. Vers 1933–1934 une polémique est apparue entre Jeffreys et R.A. Fisher, le statisticien le plus renommé de l'époque. Avant cette rencontre, les travaux de Jeffreys se penchaient sur l'usage des probabilités en tant qu'outil inferentiel scientifique ainsi que sur l'utilisation de la théorie des erreurs en astronomie et séismologie. Après cette rencontre, Jeffreys a produit la "Theory of Probability", founissant ainsi un traitement complet de la statistique bayesienne.

[*Received October 2003, accepted November 2004*]