



School of Social Sciences  
Economics Division  
University of Southampton  
Southampton SO17 1BJ  
UK

## **Discussion Papers in Economics and Econometrics**

**THE STATISTICAL EDUCATION OF  
HAROLD JEFFREYS**

by

John Aldrich

No. 0309

**This paper is available on our website  
[http://www.socsci.soton.ac.uk/economics/Research/  
Discussion\\_Papers](http://www.socsci.soton.ac.uk/economics/Research/Discussion_Papers)**

# The Statistical Education of Harold Jeffreys

John Aldrich  
Economics Division  
School of Social Sciences  
University of Southampton  
Southampton  
SO17 1BJ  
UK

Fax (0)(+44) 23 80593858  
e-mail: [john.aldrich@soton.ac.uk](mailto:john.aldrich@soton.ac.uk)

## Abstract

The paper considers the statistical work of the applied mathematician Harold Jeffreys. In 1933-4 Jeffreys had a controversy with R. A. Fisher, the leading statistician of the time. Before this encounter Jeffreys had been developing probability as the basis for scientific inference and using the theory of errors to astronomy and seismology; he was also starting to rework the *theory* of errors. After the encounter came the *Theory of Probability*, a full-scale treatment of Bayesian statistics.

October 2003

# 1 Introduction

In 1931 the Cambridge applied mathematician and physicist Harold Jeffreys F.R.S. published *Scientific Inference*, a book treating “the chief guiding principle of both scientific and everyday knowledge: that it is possible to learn from experience.” Philosophers and logicians had studied the principle but amongst physicists “it hardly seems to be noticed that such a principle exists” (p. vii). Eight years later in the *Theory of Probability* Jeffreys noticed a third group—“modern statisticians”. How he encountered them and matched their “extensive mathematical techniques” with his own Bayesian methods is the subject of this paper.

*Scientific Inference* was a work on the philosophy of science though it did something to apply the theory of probability to the treatment of measurement errors. (By “the theory of probability” Jeffreys meant a theory of inductive inference founded on the principle of inverse probability—in modern terms, Bayesian inference.) Jeffreys’s *statistical* education began in 1933-4 with a controversy with Fisher, the pre-eminent modern statistician. Afterwards in the *Probability* (p. vi) Jeffreys wrote, “I must offer my warmest thanks to Professor R. A. Fisher and Dr. J. Wishart [the Cambridge face of modern statistics] for their kindness in answering numerous questions from a not very docile pupil.” The pupil inevitably failed his teachers for he never wanted to be a modern statistician. Fisher used to say “He makes a logical mistake on the first page which invalidates all the 395 formulae in his book.” (from Box (1978, p. 441))

Section 2 presents Jeffreys before the encounter, developing probability as the basis for scientific inference, using the theory of errors in astronomy and seismology and starting to rework the *theory* of errors. Sections 3 and 4 consider the controversy with Fisher and how Jeffreys came out of it. What the *Probability* did with modern statistics is the topic for Section 5, while Section 6 examines what modern statistics made of the *Probability*.

## 2 Probability, seismology and errors

Harold Jeffreys (1891-1989) spent his life—apart from 1917-22 at the Meteorological Office—in Cambridge applying mathematics to celestial mechanics, fluid dynamics, meteorology, geophysics and probability.<sup>1</sup> He arrived as an undergraduate in 1910 and, though he retired from the Plumian Professorship of Astronomy and Experimental Philosophy in 1958, he went on working into the 1980s and his 90s. Cambridge influenced both his intellectual orientation and the details of his research. The writings of George Darwin inspired his approach to physical problems, while Jeans and Eddington were living influences, the latter perhaps provoking more than inspiring. Influences specific to probability are noted in §2.1 but Whitehead and Russell’s *Principia Mathematica* was an inspiration for the whole project and there were pure mathematicians Jeffreys could go to for help with the problems the project generated.

There are several strands to Jeffreys’s work as a physicist. He produced mathematical models of the solar system and earth; his great achievement here was an inference from one of his models, “the definite statement in 1926 that the Earth’s core is liquid”—see Spall (1980). *The Earth: its Origin, History and Physical Constitution* (1924) expounded the models but also treated many other matters including empirical seismology. Jeffreys’s interest in this field would lead to his work in statistics. His first seismological publication was in 1923.

Jeffreys also published philosophy of science which he saw as a continuation of his more mainstream scientific research. Thus he responded to the theory of relativity by considering what makes a scientific theory acceptable and by doing empirical work. Both involved probability—the first more immediately for Jeffreys applied the theory of errors long before he

---

<sup>1</sup>Cook’s (1991) memoir is particularly strong on Jeffreys’s career in physics. Howie (2002) provides a rich account of Jeffreys’s personal and intellectual life, especially in the 1920s. Both have bibliographies. There is a smaller annotated bibliography and sketch of Jeffreys’s career in Aldrich (2003). Good (1980) and Geisser (1980) give brief accounts of Jeffreys’s work in probability.

brought it within the orbit of his theory of probability.

## 2.1 Probability and scientific inference

In his long life Jeffreys often told how he became interested in probability and the principles of scientific inquiry, e.g. recalling (1963, p. 407):

I first took an interest in problems of scientific inference because my work concerned geophysics and cosmogony. It involved laws derived from observations on a laboratory scale and over intervals of time ranging from minutes to 2000 years, which were applied to distances of over 6000 km. for the centre of the earth and  $10^9$  km. for the solar system, and to times over  $10^9$  years. A usual criticism was that such extrapolation was speculative and unverifiable.

The same thought appears on page 1 of *The Earth*.

Jeffreys also recalled (again 1963, p. 407) being impressed by Karl Pearson's *Grammar of Science*: "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 read Pearson in 1914 and there is no understating the importance of this "attitude" for Jeffreys; it was the central dogma of his work on probability. But Jeffreys's early publications reflected rather the influence of contemporary Cambridge thought on probability and induction. His co-author Dorothy Wrinch (1894-1976), like the better-known C. D. Broad and J. M. Keynes, had attended the lectures of W. E. Johnson.<sup>2</sup>

The Wrinch-Jeffreys project—which Jeffreys continued alone—was to find a rational basis for what scientists *do*. They defend an axiom with the appeal, "If then we wish to retain the

---

<sup>2</sup>See Howie (pp. 85ff) for Wrinch and her collaboration with Jeffreys and Zabell (1982), von Wright (1959) and O'Donnell (1989) for Johnson, Broad and Keynes respectively.

customary applications of the theory (and this seems desirable at any cost) we must assume that axiom 2 [property 2 below] is correct.” (1919, p. 722.) Gauss’s Bayesian argument for least squares, which was taught in Cambridge,.could have been such an application.

Jeffreys’s memories of early influences seem to have been fixed in the 1930s, when he first wrote about them. In 1937 he (Bennett (p. 162)) 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.” Jeffreys’s disagreement with Eddington over the basis for relativity theory was serious but the first Wrinch-Jeffreys paper, “On some aspects of the theory of probability” (1919), predates it. “Aspects” treats alternative conceptions of probability, the classical, frequentist and logical as they would be called today. It has only two recent references. The paper drew upon the mathematical technique of “Approximations in the theory of probability” (1919) by Bromwich, a fellow of Jeffreys’s college, while it complemented Broad’s “The relation between induction and probability” (1918), as Cambridge probability for scientists instead of for philosophers.

Wrinch & Jeffreys (1919, p. 715) look back to De Morgan and Jevons; they had treated probability as “comprehensible without any definition, and perhaps, indefinable, satisfying certain definite laws the logical basis of which is not yet clear.” Probability (pp. 724-5) is an extension of classical logic to cases where premisses do not establish conclusions with absolute certainty. Wrinch and Jeffreys start at a more basic level than De Morgan and Jevons and indeed (pp. 720-1) prove one law they had “regarded as axiomatic”. They are also more basic than Broad whose “principles of probability accepted as self-evident” (p. 393) were the addition and multiplication laws for the probability of  $x$  given  $h$ , a “convention” that the probability of the certain event is unity and the “principle of sufficient reason”, as Wrinch and Jeffreys called it.

In the Wrinch and Jeffreys system numbers are assigned to combinations of propositions and data. Their system rests on a primitive notion of orderable combinations as “Property 2”—in all there are 4—(p. 720) will illustrate:

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.

This axiom “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.” Wrinch and Jeffreys found the principle unproblematic but Keynes’s *Treatise on Probability* (1922) has a chapter on “the paradoxical and even contradictory conclusions” to which the principle leads. Keynes (p. 48) reviewed objections to the principle such 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. Jeffreys (1922, p. 132) would remark 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 (p. 725) put their probability apparatus 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  $n$ ?

For Jevons and Pearson this “problem of sampling induction” was the paradigm scientific inference problem. It was also a problem in statistical inference, although statisticians like

Yule (1911, chapter xiii) favoured the variant, where the number of balls is infinite and the proportion of white balls is to be determined. Wrinch and Jeffreys do not refer to their writings. Yule was a member of Jeffreys's college from 1913 and a fellow from 1922 but there are no reports of them talking statistics before the '30s.

Wrinch & Jeffreys review the results for the case when  $f(n)$ , the prior probability of any particular number of white balls, is "independent of  $n$ " (uniform prior), adding, however

It is however very rarely, if ever, possible to assume, on the data available before the sample is taken, that  $f(n)$  is independent of  $n$ , and cases where it has other values are much more interesting.

They do not consider any such cases in detail for their main object is to show "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 derived a new and striking result from the uniform prior scheme: if  $q$  is 0, the probability that *all* the balls in the bag are white is  $(p+1)/(m+1)$ . 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 Broad's "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".

"Aspects" fixed the pattern of Jeffreys's probability in what it denied. It rejected Venn's (1888) notion of probability as a limiting frequency: "the definitions offered either implicitly

involve the very notion they are meant to avoid, or else make assumptions which are actually erroneous". Venn's formulation depends on taking the limit of a ratio of terms in an infinite sequence and "serious difficulties" present themselves (p. 716):

The existence of a probability on this theory requires that a limit shall exist to which a certain ratio tends in the long run; and one is led to ask what the evidence is for the existence of such a limit.

Jeffreys returned to the criticism of the Venn limit in *Scientific Inference* and even in the *Probability*, though there he concentrated on its "modern exponent"—von Mises. Jeffreys *never* considered the limit conception mathematically viable.

Apart from its technical deficiencies Venn's probability was too restrictive for the needs of science:

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.

(p. 716)

Jeffreys had a paper on planetary encounters (1916) although it contained no probability arguments.

There was an essential requirement that Venn's probability failed to satisfy (p. 731):

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 primacy Wrinch and Jeffreys gave to the principle that induction be based on probability.

The second Wrinch & Jeffreys paper with probability arguments, “On certain fundamental principles of scientific inquiry”, (in two parts, 1921/3) confronted the following dilemma:

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*. (1923, p. 368)

Surprisingly, given their former endorsement of Broad’s conclusion, Wrinch and Jeffreys accept the physicists’ belief. They 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. This simplicity principle became a permanent part of Jeffreys’s thinking.

The Wrinch-Jeffreys project was concerned with identifying the probability *implicit* in the practice of science. It did not engage with the probability already explicit—e.g. in gas theory, or in the combination of observations. Jeffreys used its probability formalism only once in his scientific writing of the 20s—to make the case for the tidal theory of the earth’s formation: the conclusion—translated into ordinary language—was:

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. (*The Earth*, p. 259)

## 2.2 Least squares & observational seismology

When Jeffreys started empirical research the theory of errors was a mess, institutionally and intellectually. Contributions would appear in astronomy, meteorology or general science journals and the prevailing analysis was a confusing medley of prior-less Bayesian analysis and frequentist analysis; see Aldrich (1997). The astronomer Meldrum Stewart (1920, p. 218) complained “it would be difficult to find a subject where confusion of thought and loose reasoning are more prevalent.” The subject was open to all-comers and contributions seemed no more cumulative than solutions to newspaper puzzles. Indeed Pearson (1901) had published Stewart’s contribution twenty years before in the same journal. In 1931—see §2.3 below—Jeffreys published a Bayesian treatment of the mean of the normal distribution with unknown dispersion unaware of Burnside’s 1923 effort.

In Cambridge the authorities were astronomers. Jeffreys attended Eddington’s lectures in 1913—see Howie (p. 162)—while Fisher probably attended Stratton’s lectures in 1911. Unfortunately the best guide to what was taught in Cambridge, Brunt’s *Combination of Observations* (1917), is based on Stratton’s lectures with only touches of Eddington. The subject had a traditional core though Brunt—following Stratton presumably—covered some of Karl Pearson’s work. Pearson was the most dynamic figure in what would now be called statistics but biometry was his main interest and his papers usually appeared in the obscure *Biometrika*. ‘Statistics’ at the time stood primarily for economic and social statistics, a study overseen by the Royal Statistical Society. Its journal had little to interest physical scientists, though theoretical work by Edgeworth, Bowley or Yule appeared occasionally. The arrival of R. A. Fisher forced a re-grouping for he did not fit comfortably into any group. “Modern statisticians”,

Jeffreys's phrase of 1939, matched the new reality.

Jeffreys's use of the theory of errors was not especially innovative. The objective of the astronomical investigation (1916a) was to test Einstein's theory but the use of weighted least squares to find the unknowns in a small number of dynamic equations went back to Gauss. Jeffreys's move into seismology did not involve the change of culture a move into biometry would have done, for biometry had become a separate discipline; see Aldrich (2002a). Like the work in astronomy, that in seismology was based on least squares, though there were some special features, e.g. the data was rougher and there were often many more normal equations to solve—see Jeffreys & Bullen (1935, p. 4). In Jeffreys's least squares work the study of residuals was important—to determine whether the fitted relationship was adequate (see e.g. his (1926, p. 388) ) or to detect outliers. In Wrinch & Jeffreys (1923) the velocities of  $P$  and  $S$  waves (from 5 and 4 observations respectively) are calculated using least squares. Happily they report, “All the residuals are ... within the possible errors of measurement. Accordingly our hypothesis that the observations refer to two waves spreading out with uniform velocity in a homogeneous medium is in close agreement with the facts.” (p. 16).

In the 1920s Jeffreys used statistical methods without reflecting much upon them. An exception was the thoughtful account of periodogram analysis in *The Earth's* appendix on “empirical periodicities”—extended in the 1929 edition. He did not treat the topic using the theory of probability and indeed there was an implicit conflict when he (p. 270) presented the usual significance test for the “reality” of an amplitude; he returned to the topic in his (1935a). His main resource for error *theory* was Whittaker & Robinson's *Calculus of Observations* (1924). This was a wonderful compendium of ideas and methods but it had no single viewpoint: Gauss's Bayesian and non-Bayesian “proofs” of least squares follow one another (pp. 215-226). It had some coverage of the biometric school, though Greenwood

(1924) thought this was not up to date. The very recent work of Fisher (1922, 1922a) on statistical inference and least squares/regression was not mentioned.

### 2.3 The theory of errors: needing “some modification”

*Scientific Inference* (1931) returned to the Wrinch and Jeffreys topics—probability, geometry and dynamics. Jeffreys later told Fisher that he “should have written it in 1923, when Einstein was still news.” (Bennett p. 170) There was some probability news: a “non-quantitative simplicity postulate” (pp. 191-7) moved the sampling problem forward by proposing a solution to the Broad puzzle—“ $f(r)$  ... should be taken to be inversely proportional to  $r(n - r)$ ”—and there was a chapter on “errors”. For Jeffreys the sampling problem was primarily of philosophical interest but the problem of errors was not. Still his conviction that “the usual presentation of the theory of errors needed some modification” (1932a, p. 48) seemed to come more from reflecting on Whittaker & Robinson than from seismologising.

The principal modification (1931, pp. 66-70)—another was a less than happy criticism of Gauss—involved inference to the true value,  $x$ , when the errors follow the “normal law” with the precision *unknown*; Jeffreys used the quantity  $h = \frac{1}{\sigma\sqrt{2}}$ , like many error theory writers. The modification might almost be called the *usual* one for the “usual presentation” was repeatedly extended to accommodate unknown precision—most recently and most locally, by the algebraist W. P. Burnside writing in 1923 in the Cambridge mathematicians’ journal.<sup>3</sup> Fisher (1923) had replied with a criticism of Burnside’s Bayesian approach and a reference to ‘Student’ (1908) where the job was done properly.

Jeffreys (1931, 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

---

<sup>3</sup>Pfanzagl & Sheynin (1996) review these efforts.

the position is “different with regard to  $h$ ”. Jeffreys treated the parameters  $x$  and  $h$  separately, taking it for granted that they are prior independent. A new element in his treatment was the recognition that the argument *might* be based on a second measure,  $\sigma$  the “standard error” as he called it.<sup>4</sup>

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  $\sigma$  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  $\sigma$ , 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.

The “ambiguity” is presumably that would occur if a uniform prior were given to  $h$  and  $\sigma$ . With  $P(x, h)dh$  proportional to  $dh/h$  Jeffreys derives the “probability of a value of  $x$  in the range  $dx$ , irrespective of  $h$ .<sup>5</sup> (p. 69). The resulting non-normal form (see §5.3 below) was not new in Bayesian analysis but usually the joint prior was taken to be uniform—in whatever parameterisation was adopted—without any discussion.

“On the Theory of Errors and Least Squares” (1932a) extended the inverse argument to normal regression, as Fisher (1922a) had extended ‘Student’s direct argument.<sup>5</sup> Jeffreys

---

<sup>4</sup>The terminologies and notations of least squares and biometry/statistics are surveyed in Aldrich (2003).

<sup>5</sup>Aldrich (2000a) discusses Fisher’s regression analysis.

also 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.” (1934, p. 530). This second way could also be used in significance testing—see §5.2 below.

Jeffreys (1932a, p. 48) showed that only one prior,  $dh/h$ , 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.)

The argument was summarised by Fisher (1933, p. 344) before he attacked it (§3.2 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 the  $t$  distribution. There was a more pressing problem arising in his seismological research (1931a).

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 ... (1932, p. 78)

The 1932 paper was called “An alternative to the rejection of observations”—alternative to the use of a criterion such as Chauvenet’s—and Jeffreys treated the observations as coming from a mixture of normals. He was apparently unaware that Pearson (1894) had investigated the mixture set-up and proposed the method of moments for it. For Jeffreys the problem combined features of the normal mean and sampling induction problems and his prior for the two means, two dispersions and mixture parameter reflected this.

Jeffreys was not to go on quietly developing methods for reducing observations for he was deflected by modern statistics. The encounter would not have held much promise. The “statistical theory of probability” was Jeffreys’s name for Venn’s theory and for the author of *Scientific Inference* (pp. 33-4) statistics was about taking 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.”

### 3 The encounter with Fisher

R. A. Fisher F.R.S. (1890-1962) also learnt the theory of errors from Cambridge astronomers but, though he contributed to the journals of the Royal, Astronomical, Meteorological and Cambridge Philosophical Societies, his interests in genetics and agricultural statistics removed him from Jeffreys’s world.<sup>6</sup> Howie (p. 226) traces their differences on probability to their experience as researchers in genetics and seismology. I think biometry and Cambridge philosophy

---

<sup>6</sup>For some of large literature on Fisher’s work see the bibliography in Aldrich ((2003b)).

were bigger influences—see Aldrich (2003c)—but, whatever the cause, the two were miles apart. 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.

### 3.1 The practical worker and the mathematician

In his review of Keynes’s *Treatise* Jeffreys (1922) criticised its “excessive caution”. In *his*, Fisher (1922/3, p. 46) commented:

Dr. Jeffries [sic] gives such an indulgent account [of the book] 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 unacceptable. “To the statistician probability appears simply as the ratio which a part bears to the whole of a (usually infinite) population of possibilities.”

Fisher was noticed in the section of *Scientific Inference* on the statistical theory of probability: “R. A. Fisher [(1922)] with what looks like the courage of despair, says that in a ‘hypothetical infinite population’ the ratio is perfectly definite.” (p. 220n) Jeffreys marvelled that a mathematician of Fisher’s quality could be so inept.<sup>7</sup> The foundations of probability did not trouble Fisher. To persuade Burnside that the ideas which had “grown up in the minds of practical statisticians” were sound Fisher added a “prefatory note” to the “Theory

---

<sup>7</sup>Jeffreys (1933, p. 533) : “That a mathematician of Dr. Fisher’s ability should commit himself to the statement that the ratio of two infinite numbers has an exact value can only be regarded as astonishing.”

of Statistical Estimation” (1925, p. 700) and then wrote no more on the subject for thirty years—until *Statistical Methods and Scientific Inference*. Meanwhile he produced methods and an inference theory that did not depend on psychological notions.

Fisher did not just avoid Bayesian ideas; he (1921, p. 24) campaigned against them:

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.

He (e.g. 1922, p. 326) would refer to the “baseless character of the assumptions” made under the names inverse probability and Bayes’ theorem and the “decisive criticism to which they have been exposed.”<sup>8</sup>

In 1922 Fisher considered  $p$  the probability of a success in the binomial distribution—the infinite version of the Wrinch-Jeffreys problem. He (pp. 324-5) argues against Bayes “postulate”, i.e. that it is reasonable to assume that the a priori distribution of the parameter  $p$  is uniform, 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  $\theta$ , 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 §2.1 above.

Jeffreys wrote to Fisher for a copy of this paper—the only evidence of direct communication between them before 1933—but, though he referred to the appeal to the “hypothetical infinite

---

<sup>8</sup>For more details see Aldrich (1997 and 2000).

population”, he did not answer the “indeterminacy” argument fully until his 1946 paper on “An invariant form for the prior probability in estimation problems.” (§6 below)

For Fisher, there was a determinacy corresponding to the “indeterminacy”: the inverse argument produces a “perfectly definite value” for the probability distribution of the parameters *when* there is “a super-population of known specification.” (1930, pp. 530-1) The fiducial argument appears in “Inverse probability” (1930) as a *second* way of attaching probabilities to parameter values. The two ways were compared in the paper, as they were in his next on the subject—his response to Jeffreys.

### 3.2 The controversy

After some preliminaries in the *Proceedings of the Cambridge Philosophical Society* the controversy took off in the *Proceedings of the Royal Society*—home journals for them both as Fisher was in exile from the main statistics journals.<sup>9</sup> A paper by Haldane (1932) on the sampling problem set it going. Like his earlier paper on estimating linkage (1919), this used inverse probability. Haldane did not mention Jeffreys, presumably because he did not know his work, but he more than mentioned Fisher; he interpreted maximum likelihood theory as inverse theory, with minus the second derivative of the log likelihood as the reciprocal of the variance of the large sample normal approximation to the posterior, a delicious circularity as Fisher (1922) had been correcting Pearson’s confused Bayesian reasoning from the 1890s; see Aldrich (1997).

Fisher (1932, p. 260) was exasperated: “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.” Fisher (1921) had introduced the term “likelihood” when Pearson’s

---

<sup>9</sup>Lane (1980) and especially Howie (ch. 5) use the controversy to highlight differences in each party’s conceptions of probability.

team (Soper et al. (1917)) ‘found’ the same tacit assumption. Fisher (1932, 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 to this line of argument 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 (1933) also responded negatively to Haldane, mainly to an “absurdity” Haldane found when using a uniform prior.<sup>10</sup> But then in passing he (p. 86) threw in a “fundamental objection” to Fisher’s position:

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.

---

<sup>10</sup>Jeffreys was more positive in the *Probability* (1939, p. 114) particularly about the uniform prior for the logarithm of the Poisson parameter.

Jeffreys seems to have misunderstood this particular statistician but the final sentence brings out the depth of their disagreement.

Haldane and his problem now disappear for meanwhile Fisher (1933) was criticising Jeffreys's least squares paper (§2.3 above). Fisher focussed on the prior for  $h$  and, by ignoring the  $t$  work, put back Jeffreys's education. Having learnt of 'Student's' work from Wishart, Jeffreys first wrote about this important chapter of modern statistics in 1937 (see §5.3 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 was not what Jeffreys had in mind but for Fisher it was the only meaningful question and it gave him an opportunity to elaborate his comparison of fiducial and super-population Bayesian reasoning.

In 1922 Fisher proclaimed the futility of the exercise of evolving "a vitally important piece of knowledge ... out of an assumption of complete ignorance". Now he (pp. 343-4) 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 with 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 §2.3 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  $\sigma$  (replacing  $h$  for the “convenience” of statisticians). Applying the fiducial argument to the mean, or to a regression coefficient, involved rather more than transposing and the normal mean was given the fiducial treatment in his (1935b) while regression waited until the posthumous edition of the *Statistical Methods* (1970).

Fisher (1933, p. 347) starts from the observation that, as the ratio  $s/\sigma$  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  $s/\sigma$  will be exceeded in random samples.” He proceeds to invert statements about  $s$  so they become statements about  $\sigma$ . Thus a statement of the form

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

is equivalent to the inequality,

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

since for any value of the probability chosen, the corresponding values of  $s$  and  $\sigma$  increase together from 0 to  $\infty$ .

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 a probability statement like that 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.

Another interpretation is that the fiducial argument can be supported by an inverse argument with this prior. Jeffreys would make such a case in the *Probability*—see §5.3 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.

It was also the case that

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 (1933) returned to first principles, 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.

Under fire Jeffreys 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 (1933, 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.”

In his counter-attack Jeffreys (p. 532) made a general point about averaging as it appeared in Fisher’s argument. It missed the point of Fisher’s argument but it was important because it led on to an argument which would run and run (see §5.2 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 appli-

cation and not the ones 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 letters between them are in Bennett (pp. 149-161). 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 §5.1 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) concludes his account, “Nothing was settled by the exchange. Neither scientist seems to have convinced the other of anything.” They left with their convictions reinforced and their sense of authority intact. Jeffreys went away sharper. The only criticism that moved him moved him to a more extreme position. He was also forced to go beyond the unacceptable premisses and examine frequentist reasoning. Jeffreys was drawn into Statistics, although Fisher might have opened the window a little wider.

## 4 “Shaking hands”

“Some tests of significance treated by the theory of probability” (1935a), Jeffreys’s next probability paper, took the project in a new direction. It set out from a doubt about “the usual procedure” of judging a difference significant if “it exceeds a certain rather arbitrary multiple of the standard error”. (See §5.2 below.) The reader would not know that Jeffreys had been

arguing with the master significance tester—the references are Brunt for correlation and Ramsey for ethically neutral outcomes. Fisher was also missing from the next instalment (1936). Elsewhere there were signs that the Fisher interruption would come to something.

In December 1934 Fisher read “The logic of inductive inference” to the RSS. This was a better summary of “some of the theoretical researches with which I have been associated” than anything he fired at Jeffreys. Jeffreys made a written contribution to the discussion.<sup>11</sup> Predictably they disagreed about probability. Fisher was developing the view (1936, p. 247) that Bayes framed the valid super-population argument and deserved praise for *not* publishing the invalid argument, unlike the rash Laplace who had no appreciation of “Bayes’ scientific caution”. Jeffreys (1935, p. 70) asked, “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.” Fisher’s (1935, p. 80) reference to Bayes did not convince—see Jeffreys (1939, p. 334).

Jeffreys (1935, 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

---

<sup>11</sup>It was not a happy meeting. Fisher (1935, p. 39) began by referring to the “invitation of your Council to address this Society.” Later he (p. 76) concluded from “the acerbity ... with which the customary vote of thanks was moved” that “some at least of the Society’s authorities on matters theoretical viewed these developments with disfavour.”

interprets it as a posterior probability.

However there was a more positive shading to some of the 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”. Jeffreys also described how *he* would treat one of Fisher’s problems—significance tests for contingency tables.

There was already a basis for an opening to Fisher. As early as February 1934—see (Bennett p. 150)—Jeffreys was saying, “My quarrel is not with you, but with Eddington and similar people.” Fisher’s quarrel was with Karl Pearson and later with Neyman and in these quarrels Jeffreys and Fisher could see each other as allies. They remained professed opponents but they came to an accommodation, symbolised by a story in Box (p. 442): “the feud between Jeffreys and Fisher ... ended on the day they both went to hear Eddington talk on the nature of scientific inference, and were so horrified that they shook hands and promised not to write any more rude things about each other.”

When their correspondence resumes in 1937 (see Bennett pp. 162-3) they are discussing scientific inference. In a piece in *Nature* (1937, p. 1004) Jeffreys criticised the a priori view of physical laws advanced by some physicists. Against this his own system of induction and the systems given by Fisher and “other statisticians” are all satisfactory in that

- (1) they do not treat any hypothesis as a priori certain; (2) they provide methods of choosing between hypotheses by means of observations; (3) they give estimates of the parameters involved in any hypothesis consistent with the observations;
- (4) they recognise that decisions made by them will sometimes be wrong, and give estimates of, alternatively, what degree of confidence we can reasonably attach to a

decision, or how often, supposing that we use the rules regularly, we may expect the decisions to be right; (5) they all contain provision for correcting wrong decisions when observations become capable of giving further tests, thus recognizing that scientific progress is by successive approximation.

Jeffreys and Fisher appeared to shake hands over Karl Pearson too. In July 1937 Jeffreys mentioned Pearson’s “stickiness” towards maximum likelihood (see Bennett (p. 164)) and proposed a paper for Fisher’s *Annals of Eugenics*. 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 (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 Fisher’s “idea”. In his editor’s 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”. Jeffreys (p. 147) justified this with an inverse perspective on Fisher’s theory of estimation.

Yet, while they could agree about the method of moments, they could not really agree about Pearson. Jeffreys admired the *Grammar of Science* and often said so: it “remains the outstanding general work on scientific method.” (1939, p. 313)<sup>12</sup> He also shared Pearson’s distrust of the normal distribution. All Fisher granted Pearson was the  $\chi^2$  test.

---

<sup>12</sup>However, when Jeffreys (1939, p. 102) discussed Pearson’s empirical view of probability and the experiential basis of the Bayesian method, he seemed almost disbelieving: the use of uniform priors “was based on the observed fact that sampling ratios have been about uniformly distributed.”

## 5 The *Probability* and “modern statistics”

In September 1938 Jeffreys told Fisher, “I have been wishing for some time that a public benefactor would subsidize [the publisher] to scrap *Scientific Inference* and give me a chance of writing something up to date”—see (Bennett p. 170). The following year the *Probability* came out. Jeffreys’s papers, like instalments of a serial, required only the story so far. The *Probability* was very different, a hugely ambitious work that developed the part of *Scientific Inference* not about dynamics or geometry to encompass theory, methods *and* alternative standpoints. The philosophy of modern statistics was criticised in the chapter on “frequency definitions and direct methods” but its technical contributions were exploited throughout. Fisher’s methods were covered but so too were the “Pearson laws”. Applications to physics were prominent but the argument was followed wherever it led, e.g. to randomization in experiments.

My account of the *Probability* concentrates on its relation to the “modern statistics” of the day.<sup>13</sup> I discuss the principles used to organise the material on “estimation” and “significance tests” and how they cut across those of modern statistics. I also consider two confrontations with Fisher: the fiducial argument and averaging over the sample space—two and three in Jeffreys’s (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

---

<sup>13</sup>Lindley (1986) provides a modern subjectivist’s “re-reading”. Lee’s (2001) detailed list of the book’s contents, albeit of the final edition, conveys very well the scope of the book.

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 (p. 162.))

## 5.1 “Problems of estimation”

Like a modern statistician, 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) This section and the next consider Jeffreys's treatment of these problems.

Chapter III, the main chapter on estimation (there is another on "approximate methods and simplifications"), lays down rules for choosing a prior and then derives posterior distribu-

tions 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  $P(q | p) = 1$ .

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  $\infty$ , 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 that had occurred in the course of his exchange with Fisher. Jeffreys’s “estimation” is not the estimation of modern statistics, or rather the *estimations* for, while he was learning modern statistics *it* was falling apart. Estimation in Fisher (1922 and -25) involved the efficient extraction of information about a parameter from a sample. Neyman (1935, p. 75) preferred to base estimation 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 shadows Fisherian extraction with his own version of concepts like sufficiency—see Aldrich (2002, §3)—but he makes no fuss about it. In the chapter on “approximate methods and simplifications” he portrays 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. More of a departure is the suggestion—also in Jeffreys (1938a)—that for cases where there are no sufficient statistics and maximum likelihood is “very laborious” minimum  $\chi^2$  could be used. The payoff may be an approximation to the Bayesian solution but the headline is “an approximation to maximum likelihood”.

Jeffreys has nothing systematic to say about what is to be done with the “probability distribution” when it is found. 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.” The most important practical case falls under the heading of “the method of least squares”. For each unknown quantity he (p. 129) reports the least squares value (= the mean and mode of the posterior)  $\pm$  the standard error, declaring “From the  $t$  table for 7 degrees of freedom we find that the probability of an error of 2 milligals or more in either direction ranges from about 0.08 for  $g_2$  to 0.17 for  $g_5$ . ”

This use of the  $t$  table underlines the point that, though by now Jeffreys knew statistics, he was not imprinted with the insider’s intuitive geography of where things belong. For Fisher the  $t$ -distribution, the test on a normal mean and its fiducial distribution belong together—Jeffreys disperses them and does not need to say why. He was not the insider trying to convert his fellows, like Savage or Lindley a generation later.

## 5.2 Significance testing and the $P$ integral

Jeffreys did not discuss significance testing in *Scientific Inference* and the two chapters in the *Probability* derive from his 1935a and-36 papers. The critical matter in Chapter VII on testing using direct methods was largely new.

In the preface Jeffreys (p. v) criticised physicists and statisticians for the way they treated the relationship between hypotheses and evidence. 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 the statement in Fisher's *The Design of Experiments* (1935, p. 19): "it should be noted that the null hypothesis is never proved or established, but is possibly disproved in the course of experimentation."

For Jeffreys (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.

Jeffreys had first described the technique of testing in his (1935a). Chapters V and VI of the *Probability* extend this original treatment. The preliminary sketch (chapter V, pp. 193-5) 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  $\neg q$ , 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.

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

where  $\theta$  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  $\neg q$  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  $\chi^2$  and  $t$  cases.

Jeffreys, unlike Fisher, 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. 326-7) 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.

It is interesting to look at Jeffreys's treatment of the Student case, which Fisher had made famous. The *Probability* (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  $\sim q$  will refer to one centred on a value  $\alpha$  different from 0." Jeffreys actually made three attempts at Student's problem: in the 1935a paper and in the 1st and 2nd editions of the *Probability*. In the 1st edition Jeffreys took as his "fundamental" unknowns  $\alpha$ , the mean, and the "whole variation from 0", i.e.  $\alpha^2 + \sigma^2$ , which on his principles led to the priors

$$P(q \ d\sigma | H) \propto d\sigma/\sigma$$

$$P(\sim q \ d\sigma \ d\alpha | H) \propto d\sigma d\alpha / 2\sigma^2.$$

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  $a$  the sample mean and  $s'^2$  the sample variance.

In the spirit of the second way of testing a prior (§2.3 above) 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, -48) became dissatisfied with the analysis. Jeffreys considered a further set of facts on the posterior including the requirement that when  $n \geq 2$ ,  $s' = 0$  and  $\bar{x} \neq 0$  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  $x$ :

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

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 (§3.2 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”.

### 5.3 The fiducial argument

While Jeffreys consistently criticised the reasoning behind the  $P$  integral, he changed his mind about the fiducial argument. The argument, his introduction to modern statistics (see §3.2 and §4 above), seemed at first fallacious, a natural reaction given Fisher’s (1930, -33) 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 §2.3) 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.” ( p. 326.)

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

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

where “previous knowledge”  $k$  expresses “the truth of the normal law but nothing about  $x$  and  $\sigma$ .” (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  $x - \bar{x} = z\sigma'$ , leads to

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

where  $\bar{x}$  and  $\sigma'$  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  $\bar{x}$  and  $\sigma'$  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 §2.3. 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 (1935b). 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* (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  $\sigma'$ ] 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  $\sigma$  are completely unknown.

Jeffreys (p. 311) reports that “‘Student’ called my attention to the vital word [“unique”] just after the publication [of Jeffreys (1937)] 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  $\sigma$ .” Indeed there was 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) Elsewhere he was a more orthodox Bayesian, though not wholly so for he put the sampling distribution of a statistic in the place of the likelihood. (For discussion and references 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  $\mu$  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 an embattled Fisher and then in a paper (1940) in Fisher’s journal. Fisher, in a paper (1941) that was both learned and angry, insisted that his treatment of the Behrens-Fisher problem was an extension of that of ‘Student’s’ problem. He (p. 142) mentions Jeffreys, “whose logical standpoint is very different from my own”, and then appeals 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)<sup>14</sup>

## 6 The *Probability* and “modern statisticians”

I close by considering the *Theory of Probability* as a contribution to the statistical literature, not that it was written only, or even mainly, for statisticians—see §1. Statistics is so prominent

---

<sup>14</sup>Stone (1983) and Zabell (1992) describe how Fisher reconstructed the fiducial argument at this time.

in this book, compared to *Scientific Inference*, because Jeffreys had so much catching up to do. There was not much to digest from the scientists and only posthumous publications by Ramsey and Johnson from the philosophers.

To the historian of statistics, Jeffreys's object might seem to be the reversal of Fisher's anti-Bayesian revolution of the 1920s. That was not Jeffreys's perspective: he was not writing from inside statistics, nor did he seem to realise the extent and entrenchment of earlier Bayesian work. He knew more about Edgeworth or Pearson's thoughts on probability than what followed from those thoughts! He missed the Bayesian strain in Pearson's technical writing, the papers of 1908 and -17 (Soper et al.) or the table for Bayesian analysis in *Tables* (1914).

Statistician reviewers did not see Jeffreys as a restorer either. Neyman (1940, p. 558) wrote, "The book is meant to be revolutionary ...". Irwin (1941, p. 62) had a longer perspective, "Most statisticians of to-day have been brought up on [...] direct probability methods, and to them a theory based on inverse probability throughout comes with a certain amount of strangeness and requires a little practice to follow." In 1939 there were still Bayesian survivors of the pre-Fisher era, like Bowley, but they did not look to Jeffreys for support, nor were they doing influential new work in statistical theory.

Direct theory was changing while Jeffreys was challenging it. Wilks (1941, p. 194) thought 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 were being drawn to the "extensions" of Fisher and Jeffreys, by concentrating on Fisher, did not catch their attention. The substantial additions in the 2nd (1948) and 3rd (1960) editions of the *Probability* were not responses to new statistical theory but Jeffreys's attempts to improve his earlier effort. When the Bayesian

revival came it was from immanent criticism of the extensions of Fisher.<sup>15</sup> Jeffreys (1963) discussed the new subjectivism but it did not draw any new work from him.

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 who recognised “the rational cogency of the fiducial form of argument.” (1956, p. 56) was entitled to respect. When discussing the “traditional line of thought running from Laplace to Sir Harold Jeffreys” Fisher (1956, p. 20) repeated his traditional criticism of the arbitrariness of the uniform prior (p. 17). On the other hand when Jeffreys introduced his “invariant form for the prior probability in estimation problems” (1946) it was not to meet Fisher’s objections. Fisher never said that Jeffreys influenced *him* although commentators, including Lane (p. 159) and Zabell (1992, pp. 378 & 386), have remarked on how Fisher came to resemble Jeffreys, in particular, how the later formulations of the fiducial argument incorporated points Jeffreys had made against him.

---

<sup>15</sup>Aldrich (2002 and -3a) sketch these later developments.

# References

- Aldrich, J. (1997) R. A. Fisher and the Making of Maximum Likelihood 1912-22, *Statistical Science*, **12**, 162-176.
- \_\_\_\_\_(2000) Fisher's "Inverse Probability" of 1930, *International Statistical Review*, **68**, 155-172.
- \_\_\_\_\_(2002) How Likelihood and Identification went Bayesian, *International Statistical Review*, **70**, 79-98.
- \_\_\_\_\_(2002a) R. A. Fisher and Modern Regression, unpublished manuscript.
- \_\_\_\_\_(2003) The Language of the English Biometric School, *International Statistical Review*, **71**, 109-130.
- \_\_\_\_\_(2003a) *Harold Jeffreys as a Statistician*, website  
<http://www.economics.soton.ac.uk/staff/aldrich/jeffreysweb.htm>
- \_\_\_\_\_(2003b) *A Guide to R. A. Fisher*, website  
<http://www.economics.soton.ac.uk/staff/aldrich/fisherguide/rafreader.htm>
- \_\_\_\_\_(2003c) Review of *Interpreting Probability* by David Howie, forthcoming in *History of Political Economy*.
- 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.
- Bromwich, T. J. l'A. (1919) Note on Approximations in the Theory of Probability, *Philosophical Magazine*, **38**, 231-235.
- Brunt, D. (1917) *The Combination of Observations*, Cambridge, University Press.

- Burnside, W. (1923) On Errors of Observation, *Proceedings of the Cambridge Philosophical Society*, **21**, 482-487.
- Cook, A. (1991) Sir Harold Jeffreys, *Biographical Memoirs of Fellows of the Royal Society*, **37**, 303-331.
- Fisher, R. A. (1921) On the “Probable Error’ of a Coefficient of Correlation Deduced from a Small Sample, *Metron*, **1**, 3-32. .
- \_\_\_\_\_ (1922) On the Mathematical Foundations of Theoretical Statistics, *Philosophical Transactions of the Royal Society, A*, **222**, 309-368.
- \_\_\_\_\_ (1922a) The Goodness of Fit of Regression Formulae, and the Distribution of Regression Coefficients, *Journal of the Royal Statistical Society*, **85**, 597-612.
- \_\_\_\_\_ (1922/3) Review of J. M. Keynes’s Treatise on Probability, *Eugenics Review*, **14**, 46-50.
- \_\_\_\_\_ (1923) Note on Dr Burnside’s Recent Paper on Errors of Observation, *Proceedings of the Cambridge Philosophical Society*, **21**, 655-658.
- \_\_\_\_\_ (1925) *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd.
- \_\_\_\_\_ (1925a) Theory of Statistical Estimation, *Proceedings of the Cambridge Philosophical Society*, **22**, 700-725.
- \_\_\_\_\_ (1930) Inverse Probability, *Proceedings of the Cambridge Philosophical Society*, **26**, 528-535.
- \_\_\_\_\_ (1932) Inverse Probability and the Use of Likelihood, *Proceedings of the Cambridge Philosophical Society*, **28**, 257-261.
- \_\_\_\_\_ (1933) The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters, *Proceedings of the Royal Society A*, **139**, 343-348.
- \_\_\_\_\_ (1934) Probability, Likelihood and the Quantity of Information in the Logic of

Uncertain Inference, *Proceedings of the Royal Society, A*, **146**, 1-8.

\_\_\_\_\_ (1935) *The Design of Experiments*, Edinburgh Oliver & Boyd.

\_\_\_\_\_ (1935a) The Logic of Inductive Inference, (with discussion), *Journal of the Royal Statistical Society*, **98**, 39-82.

\_\_\_\_\_ (1935b) The Fiducial Argument in Statistical Inference, *Annals of Eugenics*, **6**, 391-398.

\_\_\_\_\_ (1936) Uncertain Inference, *Proceedings of the American Academy of Arts and Science*, **71**, 245-258.

\_\_\_\_\_ (1938) Comment on H. Jeffreys Paper on Maximum Likelihood, Inverse Probability and the Method of Moments, *Annals of Eugenics*, **8**, 151.

\_\_\_\_\_ (1939) 'Student', *Annals of Eugenics*, **9**, 1-9.

\_\_\_\_\_ (1941) The Asymptotic Approach to Behrens's Integral, with Further Tables for the *d* Test of Significance, *Annals of Eugenics*, **11**, 141-172.

\_\_\_\_\_ (1956) *Statistical Methods and Scientific Inference*, Oliver & Boyd, Edinburgh.

\_\_\_\_\_ (1957) Review of *Scientific Inference* by Sir Harold Jeffreys, *Cambridge Review*, 18th May, 595-7.

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).

Greenwood, M. (1924) Review of *The Calculus of Observations* by E. T. Whittaker & G. Robinson, *Journal of the Royal Statistical Society*, **87**, 291-293.

Haldane, J. B. S. (1919) The Probable Errors of Calculated Linkage Values, and the Most Accurate Method of Determining Gametic from Certain Zygotic Series, *Journal of Genetics*,

8, 291-297.

\_\_\_\_\_ (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, *Journal of the Royal Statistical Society*, **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.

\_\_\_\_\_ (1916a) The Secular Perturbations of the Four Inner Planets, *Monthly Notices of the Royal Astronomical Society*, **77**, 112-118.

\_\_\_\_\_ (1922) The Theory of Probability [review of Keynes (1921)], *Nature*, **109**, 132-133.

\_\_\_\_\_ (1924) *The Earth*, (2nd edion -29) Cambridge, Cambridge University Press.

\_\_\_\_\_ (1931) *Scientific Inference*, Reprinted with additions 1937, Cambridge: Cambridge University Press.

\_\_\_\_\_ (1931a) The Revision of Seismological Tables, *Monthly Notices of the Royal Astronomical Society, Geophysical Supplement*, **2**, 329-348.

\_\_\_\_\_ (1932) An Alternative to the Rejection of Observations, *Proceedings of the Royal Society, A*, **137**, 78-87.

\_\_\_\_\_ (1932a) On the Theory of Errors and Least Squares, *Proceedings of the Royal Society, A*, **138**, 48-55.

\_\_\_\_\_ (1933) On the Prior Probability in the Theory of Sampling, *Proceedings of the Cambridge Philosophical Society*, **29**, 83-87.

\_\_\_\_\_ (1933a) Probability, Statistics and the Theory of Errors, *Proceedings of the Royal*

*Society, A*, **140**, 523-535.

\_\_\_\_\_ (1934) Probability and Scientific Method, *Proceedings of the Royal Society, A*, **146**, 9-16.

\_\_\_\_\_ (1935) Discussion of Fisher's Logic of Inductive Inference, *Journal of the Royal Statistical Society*, **98**, 70-72.

\_\_\_\_\_ (1935a) Some Tests of Significance Treated by the Theory of Probability, *Proceedings of the Cambridge Philosophical Society*, **31**, 203-222.

\_\_\_\_\_ (1936) Further Significance Tests, *Proceedings of the Cambridge Philosophical Society*, **32**, 416-445.

\_\_\_\_\_ (1937) On the Relation between Direct and Inverse Methods in Statistics, *Proceedings of the Royal Society, A*, **160**, 325-348.

\_\_\_\_\_ (1937a) Modern Aristotelianism: Contribution to Discussion, Supplement to *Nature*, **139**, 1004-5.

\_\_\_\_\_ (1938) Maximum Likelihood, Inverse Probability and the Method of Moments, *Annals of Eugenics*, **8**, 146-151.

\_\_\_\_\_ (1938a) The Use of Minimum  $\chi^2$  as an Approximation to the Method of Maximum Likelihood, *Proceedings of the Cambridge Philosophical Society*, **34**, 156-157.

\_\_\_\_\_ (1939) *Theory of Probability*, (2nd and 3rd editions in 1948 and -60), Oxford, University Press.

\_\_\_\_\_ (1940) Note on the Behrens-Fisher Formula, *Annals of Eugenics*, **10**, 48-51.

\_\_\_\_\_ (1942) On the Significance Tests for the Introduction of New Functions to Represent Measures, *Proceedings of the Royal Society of London, A*, **180**, 256-268.

\_\_\_\_\_ (1946) An Invariant Form for the Prior Probability in Estimation Problems, *Proceedings of the Royal Society, A*, **186**, 453-461.

- \_\_\_\_ (1963) Review of *The Foundations of Statistical Inference* by L. J. Savage and others, *Technometrics*, **5**, 407-410.
- \_\_\_\_ (1974) Fisher and Inverse Probability, *International Statistical Review*, **42**, 1-3.
- \_\_\_\_ (1980) Some General Points in Probability Theory, pp. 451-3 in Zellner (1980).
- \_\_\_\_ and K. E. Bullen (1935) Times of Transmission of Earthquake Waves, *Bur. Centr. Seism Trav. Sci.*, **11**, 3-96 & 1-106.
- 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, pp. 148-160 in Fienberg, S. E. & D. V. Hinkley (1980) (eds.) *R., A. Fisher: An Appreciation* New York, Springer.
- Lee, P. M. (2001) Detailed Table of Contents of the *Theory of Probability*, on website  
<http://www0.york.ac.uk/depts/math/histstat/lifework.htm>
- Lindley, D. V. (1957) A Statistical Paradox, *Biometrika*, **44**, 187-192.
- \_\_\_\_ (1986) On Re-reading Jeffreys, pp. 35-46 of I. S. Francis et al (eds.) *Pacific Statistical Congress*, New York: Elsevier.
- Neyman, J. (1934) On the Two Different Aspects of the Representative Method, (with discussion), *Journal of the Royal Statistical Society*, **97**, 558-625.
- \_\_\_\_ (1937) Outline of a Theory of Statistical Estimation based on the Classical Theory of Probability, *Philosophical Transactions of the Royal Society*, **236**, 333-380.
- \_\_\_\_ (1940) Review of *Theory of Probability* by Harold Jeffreys, *Journal of the American Statistical Association*, **35**, 558-559.
- Neyman, J. & E. S. Pearson (1933) On the Problem of the Most Efficient Tests of Statistical Hypotheses, *Philosophical Transactions of the Royal Society of London A*, **231**, 289-337.
- O'Donnell, R. M. (1989) *Keynes: Philosophy, Economics and Politics*, London: Macmillan.

Pearson, K. (1911) *The Grammar of Science*, third edition, (first edition 1892). London, White.

\_\_\_\_\_ (1894) Contributions to the Mathematical Theory of Evolution, *Philosophical Transactions of the Royal Society A*, **185**, 71-110.

\_\_\_\_\_ (1901) On Lines and Planes of Closest Fit to Systems of Points in Space, *Philosophical Magazine*, **2**, 559-572.

\_\_\_\_\_ (1907) On the Influence of Past Experience on Future Expectation, *Philosophical Magazine*, **13**, 365-378.

\_\_\_\_\_ (1914) *Tables for Statisticians and Biometricalians*, Cambridge: Cambridge University Press.

Pfanzagl, J. & O. Sheynin (1996) Studies in the History of Probability and Statistics XLIV A Forerunner of the *t*-Distribution, *Biometrika*, **83**, 891-898.

Soper, H. E., A. W. Young, B. M. Cave, A. Lee & K. Pearson (1917) On the Distribution of the Correlation Coefficient in Small Samples. Appendix II to the Papers of "Student" and R. A. Fisher. A Cooperative Study, *Biometrika*, **11**, 328-413.

Stewart R. M. (1920) The Adjustment of Obsevations, *Philosophical Magazine*, **40**, 217-227.

Stone, M. (1983) Fiducial Probability, pp. 81-85 of S. Kotz & N. L. Johnson (eds) *Encyclopedia of Statistical Science volume 3*, New York: Wiley.

'Student' (1908) The Probable Error of a Mean, *Biometrika*, **6**, 1-25.

\_\_\_\_\_ (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*, third edition, London, Macmillan.

Whitehead, A. N. & B. Russell (1910-13) *Principia Mathematica*, Cambridge, Cambridge

University Press.

Whittaker, E. & G. Robinson (1924) *Calculus of Observations*, Edinburgh, Blackie.

Wilks, S. S. (1941) *Theory of Probability* by Harold Jeffreys, *Biometrika*, **32**, 192-194.

Wrinch, D. & H. Jeffreys (1919) On Some Aspects of the Theory of Probability, *Philosophical Magazine*, **38**, 715-731.

\_\_\_\_\_(1921/23) On Certain Fundamental Principles of Scientific Inquiry (Two Papers), *Philosophical Magazine*, **42**, 369-390, & **45**, 368-374.

\_\_\_\_\_(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. (1. III.1.)

von Wright, G. H. (1959) Broad on Induction and Probability, reprinted in C. D. Broad (ed.) (1968) *Induction, Probability and Causation: Selected Papers*, Dordrecht, Reidel.

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.

Yule, G. U. (1911). *An Introduction to the Theory of Statistics*. Griffin, London.

Zabell, S. (1982) W. E. Johnson's "Sufficientness" Postulate, *Annals of Statistics*, **10**, 1091-1099.

\_\_\_\_\_(1992) R. A. Fisher and the Fiducial Argument, *Statistical Science*, **7**, 369-387.

Zellner, A. (ed) (1980) *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys*, Amsterdam: North-Holland.