Contributions of Sir Harold Jeffreys to Bayesian Inference*

by

Seymour Geisser

University of Minnesota

Technical Report No. 311

March 1978

Contributions of Sir Harold Jeffreys to Bayesian Inference*

by Seymour Geisser
University of Minnesota

When invited to deliver this talk my first inclination was to demur as there were few Bayesians who were not better equipped for such an assignment by virtue of being more familiar with Jeffreys' works and views. Indeed, I have never heard him speak, never met him nor in fact even know what he looks like, though I understand that he is currently in his 87th year. However, the opportunity of paying tribute to a very distinguished scientist and in the process enhancing my own knowledge of the thinking of a pioneer Bayesian soon overcame my initial hesitation.

In my own formal statistical education there was at least one glaring deficiency attributable to the then prevalent Zeitgeist (embodied in the Neyman-Pearson-Wald interpretation) which rendered the Bayesian approach, at worst, totally erroneous, at best, too restrictive and somewhere in between, outmoded. But my subsequent neglect of the writings of Jeffreys is mainly ascribable to laziness and being slightly put off by the unfamiliar logician's style in his Theory of Probability (1939), palatable mainly to those who delight in the Russell-Whitehead Principia mode. This was particularly evident in his rather distinctive notation for probability functions. Of course, these self-indulgent excuses would rarely be tolerated in a graduate student.

At any rate, it was not Jeffreys, not Savage, and certainly not de Finetti, but most curiously Fisher in his outrageous fiducial mode -- especially his book Statistical Methods and Scientific Inference (1956) who indirectly persuaded me of the Bayesian view and impelled me to develop Bayesian solutions for statistical problems. Afterwards when I

*Address to NBER-NSF Seminar on Bayesian Inference, Oct. 28, 1977, Madison, Wisconsin.

was first exposed to Jeffreys' book, Theory of Probability, I had the feeling that if one carefully scrutinized the work one could very likely find Bayesian solutions to most statistical problems -- but exasperatingly, the search often seemed to require nearly as much effort as the research. And also there was the brilliant Jimmy Savage, that all too prophetic and persuasive popularizer of Bayes, or, if one preferred pronouncements articulate, tones well-modulated, reasoning flawless, with accent impeccably British, there was his chief apostle to the heathen abroad, Dennis Lindley. But my acquaintance with the work of that 20th century master Bayesian and postulator of original ignorance (sin as some would have it) or a reasonable facsimile thereof, Sir Harold Jeffreys, was slight. So, I began again to read Jeffreys in order to try to summarize his contributions to Bayesian inference. First, one must realize that Fisher, aside fron statistics, was also a superb geneticist and applied mathematician, if not an eccentric eugenicist. Jeffreys, it turns out, is even more the Renaissance man. The field of statistics aside, his knowledge is broad and his contributions are legion -- to sciences, such as physics, astronomy, geology, cosmogony; to philosophy, especially logic; and to applied mathematics. He published several excellent books on conventional applied mathematics, elucidating mathematical methods that would serve physical scientists. He published papers on cosmogony, the origin of the universe, and then a book entitled The Earth--Its Origin, History and Constitution, (1929) so that literally his purview was all things relating to heaven and earth and especially, to quote Genesis -- "In the beginning." This engaging In-the-beginning syndrome is also reflected in his statistical work as I shall shortly discuss.

In order to appreciate and perhaps understand Jeffreys' attitude towards statistical inference, one must first discern his views on epistemology -- the theory of the nature and grounds of knowledge -- or as O. Kempthorne[1] once, crudely but succinctly, put it "the knowledge racket;" and secondly, Jeffreys' attitude towards scientific endeavor.

His book, Scientific Inference (1931), addressed the question of the nature of inference from empirical data so as to predict events that may occur in the future or retrodict experiences that were unrecorded in the past. It was Jeffreys' view that the goal of science, which is to deepen our understanding of natural processes, could best be accomplished or verified by the predictive capacity resulting from the modelling system. Clearly then, Jeffreys was philosophically a predictivist, although his applications did not always emphasize this perspective as much as one would expect.[2] I suspect he was distracted from this goal because he wanted to provide solutions for the then contemporary statistical paradigms which stressed parametric estimation, in order for his work to obtain a hearing amongst statisticians. Also advising scientists on the measurement of physical constants with an apparatus subject to error must have been of some concern to him as well as his view that physical laws were probabilistic in nature and involved these entities.

Returning to the book, Scientific Inference -- it begins with a witty and perceptive Platonic dialogue between a botanist and a logician elucidating induction and scientific law or more mundanely establishing that inference from past observations to future ones cannot be deductive. Jeffreys insisted that all scientific laws were tentative (just as Fisher insisted that all hypotheses were provisional) exhibiting uncertainty and were essentially successive approximations that never achieve

---

1. Oral communication.

2. In this regard it is curious to note that Jeffreys (1939) who discerned that Karl Pearson's Grammar of Science (1892) was inconsistent with some of his later papers, accuses Pearson of not being much influenced by his own writings as one would expect.

finality. Incidentally, for a physicist he had a deep appreciation of biological variation as can be inferred from the dialogue. He believed that it was no different in kind from variation in physics though perhaps different in degree, at a time when physicists were still enhanced by the "exactness" of their science embodied in what he called the certainty and exactness fallacies, (i.e., that scientific laws are statements made with certainty and physical measurements can be exact) two fallacies which he attacked vigorously.

He also maintained that to give any systematic account of the scientific method, it is necessary to have a conception of partial proof -- a many valued logic or a probabilistic logic where for him probability is a relationship between a set of data and a conclusion in the sense of representing degrees of reasonable belief. Probabilistic logic then must serve as the basis for the inductive argument and in fact, the inductive argument is a "logical" argument only within the context of a Bayesian framework.

It is instructive to trace the origin of Jeffreys' philosophical platform. In England in the 19th century, empiricism was on the ascendancy; experience was to be the sole guide. This empirical view was a form of idealism, that theory of knowledge which maintained nothing exists but the mind of the observer and that the external world is merely a mental construct to give oneself a convenient way of describing one's experience. Its competitor was the, then, form of realism, a theory of knowledge that held that the external world exists independent of the observer and that the function of the scientific method is to discover its properties. A special form of idealism called phenomenalism, developed by Karl Pearson (1892) and Ernst Mach (1883) asserted that nothing can be presumed to exist that cannot be reduced to a description of sensations. It requires analysis of suggested scientific laws to exhibit what they actually say about experience and if such a law refers to quantities whose values do not effect the prediction of experience, then the law should be restated so that these quantities do nôt appear in it. Although this Pearson-Mach formulation of idealism was Jeffreys' early philosophical stance, later he felt it was too stringent. In particular, he did not think that everything mentioned in a scientific law must be separately observable. His position came to be what he called critical realsim which he describes as one that maintains that inferences that go beyond the original observations are valid though uncertain as opposed to either naive realism or idealism which required that inferences about observations or parameters be made with certainty.

Given his views, Jeffreys set out to construct a logical apparat for scientific induction which is Bayesian in form. Using his approach in his books, he set forth in detail a comprehensive and normative Bayesian inferential approach for a wide variety of statistical paradigms in estimation and testing -- but scantily few for prediction -- which is surprising, considering some of his initially stated aims. It is likely, as I have previously mentioned, that he wanted to display and compare the Bayesian approach with its competitors which had paid little or no attention to prediction and that certain parameters can potentially be more than just artificial, or even conceptually meaningful constructs to physical scientists. It would take too long to catalogue his clever resolutions for these paradigms, many of which are, no doubt, quite familiar to you. Jeffreys' views were first put forth during the massive tide directed against inverse probability, initiated by the works of Fisher (1922) and carried on by Neyman (1937) -- whose preeminence and authority was such that non-Bayesian approaches became all pervasive during the 30's and 40's of this century. Almost a solitary Bayesian force during this entire period, Jeffreys managed to resuscitate Bayesianism and give it form (logical status) and substance (solutions for the pressing statistical paradigms of the day). Early on he was attacked by Fisher (1934) because his theory, involving logical degrees of rational belief, was not a frequency theory -- which Jeffreys in turn severely criticized in his books (1931, 1939) and sundry papers.

Few have engaged in polemics with Fisher and come away unscathed--and Jeffreys was no exception. An example of just such an exchange occurred some 45 years ago. It concerned the probability that a third measurement was included in the interval found by the first two--all being independently and identically distributed. Without commenting on the details--the polemics themselves are of some interest. In a heavy-handed though partially astute rejoinder to an attack by Fisher (1933), Jeffreys (1933) says:

> "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."

Fisher's response (1934), more temperate than usual, was nevertheless rapier sharp in its thrust:

> "Any defence which Jeffreys might have to offer of his omission to perform these integrations is thus lost in a polemical haze which his subsequent paragraphs do nothing to elucidate. I am not inclined to deny that the integrations reduce Jeffreys' theory to absurdity."

For Jeffreys to come off second best, irrespective of the validity of his arguments, to an acknowledged master of polemics, is really no cause for surprise or chagrin especially when he could devise such a strikingly penetrating summary of Fisher's significance testing program by poignantly heaping negative upon potential in a resounding crescendo:

> "What the use of P[the significance level] implies, therefore, is that a hypotheses that may be true may be rejected because it has not predicted observable results that have not occurred." Jeffreys (1939)

More generally Fisher (1934) accused Jeffreys of being subjective and psychological--committing the archsin of subjectivisim. Thirty years later, when this very sin suddenly turned into a God-like commandment under a new dispensation, he was accused of violating it by Savage and others. Savage (1962) labelled Jeffreys' theory as a "necessary theory", that is an objective one, and faulted it for not being subjective and then somewhat paradoxically insisted that the only objective theory possible is a subjective one. One begins to wonder at the arcane use of the terms objective and subjective and whether to call Jeffreys' arguments subjective or objective is at all relevant. For if they are subjective as Fisher would have it, then there is an enormous attempt at objectivity in providing standards for prior distributions depending on circumstances, and if basically the theory is objective (often pejoratively designated naive[3]) as Savage would have it, then the closely argued and intricate analyses for many of his canonical prior distributions would make any subjectivist envious.

Of course the single major criticism leveled at Jeffreys by the subjectivistic dispensation has been his attempts to quantify or express knowing little or ignorance, as it were. Jeffreys responds that it is not an exact quantification but some sort of workable approximation and that it is a necessary ingredient for any Bayesian view. He grants that if one has some prior information it can be taken into account. But, he did not think that this was sufficient justification for not having some prior canonical expression for approximating knowing little. To judge from his writings, his reasons were twofold, the first could be termed the "Watergate Interrogatory Syndrome". Those of you who recall those memorable hearings can hardly forget Senator Howard Baker's thundering refrain when interrogating a witness, "What did you know and when did you know it?" Presumably, by such a device a subjective assessor who had some prior informa-

3 It is possible that this disparaging label actually derives from an attempt at turning the tables on Jeffreys' (1931) characterization of certain extreme epistemologies as naive realism and naive idealism. But perhaps this imputes a much deeper meaning to the intended criticism.

tion would be driven back to the cradle or womb if necessary to reveal a time when what he knew was negligible or irrelevant to the matter at hand. His second reason was, I suppose, the public nature of science where one hoped a set of data could speak for itself and that the prior distribution was to set the Bayesian machinery in motion providing some initial neutral or impartial stance. If different individuals guess at their prior probabilities, they will ordinarily be different and the subjectivist considers this to be beyond remedy. But to Jeffreys this was the very reason to seek some canonical rule. In fact, to Savage's statement that it has proved impossible to give a satisfactory definition of the tempting expression "know nothing," Jeffreys (1963) responds "Who needs a definition." And, so it goes.

Although I do not have the time to present a catalogue of the contributions to Bayesian inference by Jeffreys, I would like to give at least a cameo presentation, extracted from Wrinch and Jeffreys (1921) and Jeffreys (1939), which in some sense reflects much of the flavor of his approach.

Recall the Bayes-Laplace rule of succession for binary events. Suppose $X_1, X_2, \ldots, X_n \ldots$ is a sequence of independent Bernoulli trials each with probability θ of being Type I say. The Bayes-Laplace assumption of a uniform prior density on θ yields $(t+1)/(n+2)$ for the chance of a Type I event on the $n+1$ trial given t Type I observations on the first n trials. If $t = n$ and n grows then the probability of a Type I on a future trial tends to 1. This was applied by many as an argument for induction of a general claim (that a particular law is always true). More than 55 years ago Jeffreys recast the problem using a finite total number of binary trials say, N where an unknown number R is of one type and $N-R$ of another. A sample of n is drawn and the number T of type I is observed with probability

$$\Pr(T=t|n,N,R) = \frac{\binom{R}{t}\binom{N-R}{n-t}}{\binom{N}{n}} \qquad t = 0,1, \ldots, \min(R,n)$$

$$= 0 \quad , \text{ elsewhere} .$$

The object now is to predict R. Assuming ignorance he posits a prior uniform probability on R,

$$\Pr(R=r|N) = (N+1)^{-1}$$

for $r = 0, 1, \ldots, N$.

Calculation of the predictive probability now yields

$$\Pr(R=r|N,\ n\ T=t) = \frac{\binom{r}{t}\binom{N-R}{n-t}}{\binom{N+1}{n+1}} \quad r = t, \ldots, N-n+t \quad .$$

$$= 0 \quad \text{elsewhere} .$$

The computation of the predictive probability that the $n+1$ observation is of type I is the expectation of the probability of this event given T, N, n, and R wrt the above predictive probability for R. This yields

$$E_R\left[\frac{R-t}{N-n}\right] = \frac{t+1}{n+2}$$

identical to the Bayes-Laplace rule of succession but with a finite horizon. In fact, this whole set-up is an observational or aparametric analogue of the original parametric model.

Now suppose the sample is wholly of Type I, i.e. $t = n$ and a general law is at issue which states that all of the N trials are of Type I or generally of one kind. Then to Jeffreys, the chance that $R = N$ should not be small if there is some real general law at stake. Calculation yields

$$P(R=N|N,\ n,\ T=n) = \frac{n+1}{N+1}$$

which is obviously small as N grows, for fixed n. This fact Jeffreys considers to be an unmitigated disaster for the Bayes-Laplace rule of succession to be a useful inductive argument. He reasons that common sense dictates that if a long series of trials were all of one kind, a feeling that this phenomenon would persist should be induced i.e. would be a true law, which would essentially be denied in the Bayes-Laplace formulation. Without giving all of the details of his argument, he revises the prior probability,

$$P(R=r|N) = \frac{1-2k}{N-1} \quad \text{for} \quad r = 2, \ldots, N-1$$

$$P(R=0|N) = P(R=N|N) = k$$

where $(N+1)^{-1} \leq k \leq 1/2$,

to attain a predictive probability more in accord with common sense

$$P(R=N|N, n, T=n) = \frac{(n+1)(N-1)k}{(n+1)(N-1)k+(N-n)(1-2k)} .$$

His next step is to posit a reasonable value for k . For $N = 2$ he notes that $(N+1)^{-1} = 1/3$. Since 1/4 is then too small he tentatively assigns. $k = 1/4 + \frac{1}{2(N+1)}$ whose values are 1/2 if $N = 1$, 5/12 if $N = 2$ etc., this yields

$$Pr(R=N|N, n, T=n) = \frac{(N+3)(n+1)}{(N+1)(n+3)}$$

so that as N grows the above tends to $\frac{n+1}{n+3}$ which he regards as entirely sensible. He recommends this solution whenever there is a serious possibility that a set of trials under consideration will all be of one type. This cameo portrait of his reasoning, at the very least, certainly refutes the charge that he was always trying to define ignorance.

If one were to present a short selected summary of Jeffreys' contributions to Bayesian inference, I believe the following would be on everybody's list.

1. He made the inductive argument a "logical" one within the context of a Bayesian framework and maintained it could only be so within this framework.

2. He made a valiant attempt to quantify lack of knowledge by giving rather clever canonical rules and conventions but was not constrained to think only in these terms.
3. He produced a normative catalogue of cogently reasoned Bayesian solutions to many conventional statistical paradigms.
4. He introduced and developed invariance considerations into the Bayesian system.
5. His devastating critiques of the various frequency theories propounded by Venn, Fisher, Neymann and others were in the words of de Finetti (1970), "closely argued and unanswerable".

In summary, Jeffreys' approach amalgamated a Bayesian system with two primitive data principles reflective of public scientific work: (1) Letting the data speak for themselves and (2) the actual units in which you choose to express your work should by and large not effect the inferenee. This is translated into so-called non-informative priors and invariance under suitable transformations. It was a rather remarkable conception, brilliantly executed, whose ultimate test is how it works in practice.

Indeed, I believe we all owe a great debt to Sir Harold Jeffreys which I personally am delighted to be able to publicly acknowledge.

References

de Finetti, B. (1970). The Theory of Probability, Vol. 2, John Wiley & Sons. (English translation, 1974) .

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics, Phil. Trans. Roy. Soc. A, 222, pp. 309-368.

Fisher, R. A. (1933). The concepts of inverse probability and fiducial probability referring to unknown parameters. Proc. Roy. Soc. A, 139, pp. 343-348.

Fisher, R. A. (1934). Probability likelihood and quantity of information in the logic of uncertain inference. Proc. Roy. Soc. A, 146, pp. 1-8.

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

Jeffreys, H. (1929). The Earth: Its Origin, History and Physical Constitution. Cambridge University Press.

Jeffreys, H. (1931). Scientific Inference. Cambridge University Press.

Jeffreys, H. (1933). Probability statistics and the theory of errors. Proc. Roy. Soc. A, 140, pp. 523-535.

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

Jeffreys, H. (1963). Review of "The Foundations of Inference" by L. J. Savage, et. al. Technometrics, pp. 407-410.

Mach, E. (1883) The Science of Mechanics, The Open Court Publishing Co. (English translation, 1960).

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. Phil. Trans. Roy. Soc. A, 236, pp. 333-380.

Pearson, K. (1892). The Grammar of Science, Walter Scott, London, Everyman's Library No. 939, 1937.

Savage, L. J. et. al. (1962). The Foundations of Statistical Inference. Metheun, London and Wiley, N. Y.

Wrinch, D. H. and Jeffreys, H. (1921). On certain fundamental principles of scientific inquiry. Phil. Mag., 42, pp. 369-390.