

R. A. Fisher on Bayes and Bayes' Theorem

John Aldrich *

Abstract. Ronald Fisher believed that “The theory of inverse probability is founded upon an error, and must be wholly rejected.” This note describes how Fisher divided responsibility for the error between Bayes and Laplace. Bayes he admired for formulating the problem, producing a solution and then *withholding* it; Laplace he blamed for promulgating the theory and for distorting the concept of probability to accommodate the theory. At the end of his life Fisher added a refinement: in the *Essay* Bayes had anticipated one of Fisher’s own fiducial arguments.

Keywords: Ronald Fisher, Thomas Bayes, Pierre-Simon Laplace, Dennis Lindley, Bayes’ Theorem, Inverse probability

1 Introduction

“R. A. Fisher was a lifelong critic of inverse probability.” “Thomas Bayes was one of R. A. Fisher’s heroes.” How can (Zabell 1989, p. 247) and (Edwards 2004, p. 34) both be right when inverse probability is only an old name for Bayesian inference? Rather easily, I suggest, for Fisher shifted responsibility for inverse probability from Bayes to Laplace and ultimately discovered an intellectual kinship with Bayes.

Ronald Fisher (1890-1962) is a critical figure in the history of Bayesian analysis by Fienberg (2006) and most students would agree with (Zabell 1989, p. 254) that Fisher and Neyman were “the two most important persons” involved in the “demise” of inverse probability in the early twentieth century. For 40 years Fisher wrote about Bayes. All he had to say about the *validity* of inverse probability, the Bayesian argument, he said at the beginning, in 1921 and -22. This is familiar ground: see Aldrich (1997) and Stigler (2005) and the references they give. Zabell (1989) has treated Fisher’s account of developments *after* Laplace but the more intriguing issue of his take on the earlier history has not received the same attention. In recent decades, as interest in Bayes has grown, Fisher’s statements about Bayes have been examined, most fully by Dale (2003), though naturally with an eye to Bayes, not Fisher. Modern works like Dale (2003), Stigler (1986), Hald (1998) and Fienberg (2006) have a very different take on Bayes and Laplace from Fisher’s. Fisher’s assessment of Laplace and, say, Hald’s are worlds apart: what Fisher knew of Laplace, which was not much, he (Fisher 1958, p. 274) thought was largely “fallacious rubbish;” while from his study of Laplace, Hald (1998) concludes that the statistical tradition is mainly a series of footnotes to Laplace!

Most of Fisher’s Bayes story was done by 1936 but in his last years Fisher changed his interpretation of Bayes. The plan of the paper is this: Section 2 introduces Bayes,

*Division of Economics, School of Social Sciences, University of Southampton, Southampton, UK, <mailto:jcal@soton.ac.uk>

Section 3 Fisher's ideas on probability, Section 4 the Laplace versus Bayes theme and Section 5 Fisher's final view of Bayes.

2 Bayes's argument

The essay by Bayes (1763) was communicated to the Royal Society after his death by Richard Price; its contents are familiar from the standard works referred to above. Fisher first wrote about the essay and the theorem in his Fisher (1921). There is an intricate back-story from Fisher (1912) proposing the "absolute criterion." This has been told several times—see Aldrich (1997) and Edwards (1997)—and fortunately is not relevant to the present study.

As (Edwards 2004, p. 35) indicates, "Bayes' theorem" for Fisher and most of his contemporaries did *not* refer to "a simple and uncontroversial theorem in conditional probabilities" but rather to "the whole inverse argument in the binomial case according to which (as we should now say) the posterior probability for a binomial parameter might be obtained from the likelihood by assuming a uniform prior probability distribution."

Fisher (1921) discusses Bayes and Bayes' theorem only to clear himself of the charge that he had used the theorem and to clarify what he *had* done: the "absolute criterion" or "the optimum" or "maximum likelihood" did *not* assume a uniform prior, or any other. (Fisher 1921, p. 24) states that Bayes "attempted to find, by observing a sample, the actual *probability* that the population value lay in any given range" adding that "Such a problem is indeterminate without knowing the statistical mechanism under which different values of p come into existence; it cannot be solved from the data supplied by a sample, or any number of samples, of the population." In his own approach "No assumption as to *a priori* probability is involved" (Fisher 1921, p. 4)—appropriately so given that "We can know nothing of the probability of hypotheses or hypothetical quantities." (Fisher 1921, p. 25)

In his "Mathematical foundations of theoretical statistics" Fisher (1922) treats Bayes' theorem as the accepted theory of estimation. He presents Bayes's argument, explains its flaw and mentions some literature; this major paper is described by Aldrich (1997) and Stigler (2005). Fisher treats the original Bayes problem of making an inference to p , the probability of success, based on x successes in n Bernoulli trials, according to Bayes and according to maximum likelihood. Fisher (1922, p. 324) lays out Bayes's problem and procedure in this way:

what is the frequency distribution of the values of p in populations which are selected by the restriction that a sample of n taken from each of them yields x successes. [...]. To render it capable of mathematical treatment, Bayes introduced the datum, that among the populations upon which the experiment was tried, those in which p lay in the range dp were equally frequent for all equal ranges dp .

Coming after his reference to a "statistical mechanism" in 1921 it is no surprise that

Fisher reads Bayes's investigation entirely in frequentist terms: the "problem" (to obtain the posterior distribution) involves frequencies of populations and so does the "datum" (the prior). This part of Fisher's Bayes story never changed: Bayes was a frequentist.

Fisher's objection to the assumption of a uniform prior was that if it were made for a different parametrisation the inference would change. (Fisher 1922, p. 324) recognised that Bayes went beyond the *axiom*, asserting equality for "all equal ranges dp ": "Bayes adds a *scholium*, the purport of which would seem to be that in the absence of all knowledge save that supplied by the sample, it is reasonable to assume this particular a priori distribution of p ." Fisher never described the scholium and when he (Fisher 1962, p. 118) next mentioned it, it was to insist that the "actual mathematics" did not depend on it. Recent commentators have treated the scholium as essential to the paper, with Edwards (1978) and Stigler (1982) arguing that its argument actually immunises the uniform prior for the probability of a success from Fisher's objection.

Pearson (1920) supplied Fisher with references to the literature on Bayes' theorem. How many he read at the time is unclear. Boole and Venn, for instance, ran and ran in his writings as redoubtable critics of the "equal distribution of ignorance" before Fisher (1956) wrote about them in detail. Zabell (1989) questions whether these critiques contain the "decisive criticism" Fisher claimed they did.

The "lifelong critic of inverse probability" was born. "The theory of inverse probability is founded upon an error, and must be wholly rejected" (Fisher 1925a, p. 10) was the verdict. Fisher first knew Bayes as the perpetrator of an error—in proposing the uniform prior for all cases where knowledge is absent, not in the "actual mathematics" for Fisher never disputed Bayes's calculations. However Bayes's two-stage experiment in which ball W is first thrown on the table and then ball O several times was an instance of the "trivial case when the population is itself a sample of super-population the specification of which is known with accuracy." (Fisher 1925a, p. 10)

3 Probability: definitions

In Fisher's Bayes story Bayes is a frequentist and Laplace a subjectivist. Fisher knew two definitions of probability, a right one and a wrong one. The right one, based on frequency, did not need elaborate examination and the wrong one did not warrant it. The wrong one, a subjective, psychological definition, was the basis for the contemporary work of Keynes and Jeffreys; see Howie (2002), Aldrich (2005) and Aldrich (2008).

Fisher exhibited the right and the wrong when he (Fisher 1922/3, p. 46) reviewed Keynes's *Treatise on Probability*:

To the statistician probability appears simply as the ratio which a part bears to the whole of a (usually infinite) population of possibilities. Mr. Keynes adopts a psychological definition. It measures the degree of rational belief to which a proposition is entitled in the light of given evidence.

Cf. the “Mathematical foundations” (Fisher 1922, p. 312) for probability as frequency in an infinite set. Apart for the odd sentence and a paragraph in (Fisher 1925b, p. 700) inclining to a limiting frequency definition, he did not write on probability until 1956.

4 Laplace versus Bayes

Laplace was there from the beginning too, or at least from Fisher (1922). The “Mathematical foundations” (Fisher 1922, p. 326) credits Laplace and Poisson with having “laid the foundations of the modern theory of statistics” although “credits” may not be quite the word as the foundations were flawed. Fisher knew there was a substantial Laplace oeuvre but was never interested enough to study it. He knew to associate Laplace with the normal distribution and with analytical developments related to cumulants (see the “historical note” in Fisher (1932)) but otherwise he seems to have known only the *Essai Philosophique sur les Probabilités*, which he had read by 1930. Surprisingly he seems to have made no use of the long chapter in Todhunter (1865). Todhunter saw Laplace as the great mover in inverse probability, like the modern commentators, including Stigler (1986), Dale (1990) and Fienberg (2006), and like Fisher.

Fisher (1930) pushed along the Bayes-Laplace story. Again Fisher had an argument he wanted to distinguish from inverse probability; the new argument—the fiducial argument—and how it came out of his earlier work are described in Aldrich (2000). Incidentally, as (Fienberg 2006, p. 14) notes, Fisher introduced the term “Bayesian” in his 1950 volume *Contributions to Mathematical Statistics* when he wrote notes on this paper and on Fisher (1921).

Fisher (1930) lays out the arguments behind the fiducial argument and inverse probability but also gives a history of the latter. A new fact and a new theme appear—the fact (Fisher 1930, p. 528) Bayes’s reluctance to publish:

Bayes, who seems to have first attempted to apply the notion of probability [...] to causes in relation to their effects, invented a theory, and evidently doubted its soundness, for he did not publish it during his life. It was posthumously published by Price, who seems to have felt no doubt of its soundness.

(Edwards 2004, p. 37) comments politely, “it is not clear on what evidence Fisher believed that Bayes withheld publication deliberately.” Price, who published the paper, made no such suggestion and Bayes’s biographers—Dale or Bellhouse—do not discuss the withholding hypothesis; it is mentioned by Stigler (Stigler 1986, p. 128–9) but he does not detect any lack of confidence in what Bayes wrote. I do not think there is any additional evidence; the conjecture is Fisher’s way of rationalising the posthumous publication, his high estimate of Bayes’s ability and low estimate of Bayes’ axiom. It seems like wishful thinking.

The new theme was the contrast between Bayes’s (very reasonable) caution and Laplace’s rashness. (Fisher 1930, p. 528) continues, “Laplace takes for granted in a

highly generalised form what Bayes tentatively wished to postulate in a special case.” Fisher also comments on Laplace’s definition of probability, called the “classical definition” today. In his (Fisher 1956, p. 15) Fisher quotes two versions of the definition. In the *Essai* version the key clause is “a certain number of equally possible cases, that is to say, to cases whose existence we are equally uncertain of” (Dale’s translation of Laplace (1995)) That this “reduces all probability to a subjective judgement” became a cardinal point in Fisher’s Bayes-Laplace story.

In 1932-4 Fisher was more engaged with inverse probability than he had ever been for he was arguing with J. B. S. Haldane and Harold Jeffreys about his own position and about the Bayesian alternative; see Howie (2002) and Aldrich (2005). These controversies, however, do not seem to have affected the Bayes story, except perhaps by giving Fisher a more exalted view of what was involved. In 1935 he had two publications dealing in part with the “logic of inductive inference.”

Inverse probability appears in chapter 1 of Fisher (1935a); Fisher believed that the book was a practical proof that the problem of induction had been solved without using Bayes’s axiom. Nevertheless he (p. 6) praised Bayes:

That he seems to have been the first man in Europe to have seen the importance of developing an exact and quantitative theory of inductive reasoning, of arguing from observational facts to the theories which might explain them, is surely a sufficient claim to a place in the history of science. But he deserves honourable remembrance for one fact also in addition to those explained by de Morgan.

The “fact” was Bayes’s readiness to *suppress* the method that De Morgan (from whom Fisher had quoted extensively) had so extolled.

The “Logic of Inductive Inference” Fisher (1935b) covered some of the same ground. In the discussion Fisher (Fisher 1935b, p. 78) re-iterated his old claim about Bayes the frequentist:

following Bayes, and, I believe, most of the early writers, but unlike Laplace, and others influenced by him in the nineteenth century, I mean by mathematical probability only that objective quality of the individual which corresponds to frequency in the population, of which the individual is spoken of as a typical member.

The novelty is the explicit contrast in the notions of probability entertained by Bayes and Laplace.

Jeffreys, a discussant, was exercised by Fisher’s super-population interpretation of Bayes’s argument: “I should be interested to know the source of Professor Fisher’s remark that in the theory of inverse probability the method was to introduce a postulate concerning the population from which the unknown population was supposed to be

drawn.” ((Jeffreys 1935, p. 70)) Fisher replied that “the procedure of Bayes is quite explicit” (Fisher 1935b, p. 80) and quoted the paragraphs from Bayes’s *Essay* (Bayes 1763, p. 385) describing the table and the throw of ball *W* and then the *n* throws of ball *O*. Fisher identified the populations in the two experiments as follows:

The casts with the ball *O* constitute a sample of events drawn from a population characterized by a certain frequency of the “happening of the event,” which is later taken to be unknown. This population is itself explicitly obtained by the previous cast of the ball *W*.

Jeffreys was not convinced and they were arguing again in 1938; see the letters in (Bennett 1990, p. 169–72).

Fisher (1936) reprised the story adding some drama. The change in attitude to Bayes’ axiom and the change in the interpretation of probability are brought together when he (Fisher 1936, p. 247) described how Bayes’ axiom had been abused:

Laplace [...] incorporated it into the foundations of his “Théorie Analytique des Probabilités,” cruelly twisting the definition of probability itself in order to accommodate the doubtful axiom. It is certain that Laplace had no appreciation of Bayes’ scientific caution. He says of Bayes, “Et il y est parvenu d’une manière fine et très ingénieuse, quoi qu’un peu embarrassé.”

The quotation is from the historical notes in the *Essai* (p. 120 in Dale’s (1994) translation). Fisher’s application of “scientific” to Bayes’s caution may seem curious when Laplace was much more of a scientist than Bayes but more of that would follow.

Twenty years later in *Statistical Methods & Scientific Inference* (1956) Fisher retold the Bayes-Laplace story in style with long quotations from their writings. The tone is more measured: Bayes is not quite the towering figure in the history of science and, instead of arraigning Laplace for “cruelty,” (Fisher 1956, p. 15) merely notes that he “needed a definition wide enough to be used in the vastly diverse applications of the *Théorie analytique*.” For Fisher the “vast diversity” of the applications was nothing to be proud of.

“Scientific caution” is emphasised with Bayes placed among those who require “sufficient objective evidence” when they make probability judgements and Laplace among those who have thought that “equality of probability may be asserted merely from the indifference of, or the absence of *differentiae* in the objective evidence.” (Fisher 1956, p. 14) Fisher then presents their definitions, noting the “difference of concept.” This was the first time that Fisher had given Bayes’s definition: “The *probability of any event* is the ratio between the value at which an expectation depending on the happening of the event ought to be computed, and the value of the thing expected upon its happening.” He (p. 14) takes this as essentially equivalent to “the limiting value of the relative frequency of success.” Fisher had been implying as much for 30 years but it is still a shock to see him write it.

5 “Bayes’ method”

Fisher’s Bayes story related how Bayes posed an important problem—solved only in the 20th century by Fisher himself—found a false solution but withheld it and how Laplace adopted the rejected solution, distorting the notion of probability and generally shunting statistics onto the wrong track. In 1959 Fisher produced a new more positive reading of Bayes for he had found something to admire in what Bayes—or rather Price—actually published! The change came about in a roundabout way as Fisher persuaded himself that he was repeating what Bayes had done two centuries earlier.

One of the objectives of *Statistical Methods & Scientific Inference* (1956) was to define the circumstances in which each of the different “forms of quantitative inference” is appropriate; it appears that “observations of different kinds may justify conclusions involving uncertainty at different levels.” In some circumstances the fiducial argument (from 1930) may be available, while in others the statistician has to be satisfied with likelihood inference (from 1921). A new form of inference in 1956 was the combination of “observations of two kinds” that separately support likelihood and fiducial inferences: the fiducial distribution supplies the “element of frequency needed to complete Bayes’ method” (p. 125). Fisher’s example involved radioactive particles and combined observations from the exponential distribution and observations from the binomial, the former contributing the fiducial distribution and the latter the likelihood function.

The new development produced no re-interpretation of Bayes; this came as Fisher thought about Dennis Lindley’s criticism in his review of the book. (Lindley 1957, p. 281) presented a counter-example to the claim that fiducial probabilities obey Bayes’ theorem and developed the point further in Lindley (1958). (Fisher 1960, p. 299) responded by charging Lindley with a “failure to recognize the problems to which the method may be correctly applied” and by giving (1962) further examples of the argument. A more unexpected consequence was a new interpretation of Bayes.

The review annoyed Fisher and he discussed what to do about it with George Barnard; see (Bennett 1990, p. 36–42) and Barnard (1987). After a while Fisher was saying, “the more I consider it, the more clearly it would appear that I have been doing almost exactly what Bayes had done in the 18th. century.” ((Bennett 1990, p. 38)) Fisher wrote this insight into the second edition of the book (Fisher 1959, pp. 17 and 127–8). The discussion of the example now (p. 128) ended with the statement that it “exhibits Bayes’ own method, replacing the billiard table by a radioactive source, as an apparatus more suitable for the 20th century.”

Portraying Bayes as “a true Fisherian” (modifying the title of Barnard (1987)) was a nice touch: it made the story better and it scored off Lindley who thought *he* was the follower of Bayes. However Fisher made no effort to justify the new interpretation or explain how it was connected with the super-population interpretation he had always given; this was, after all, a major change in Fisher’s reading of history. (Later Barnard (1987) investigated Fisher’s claim and found merit in it.) In other respects the new reading changed nothing. On the old reading Bayes had treated a “trivial case” and now with all his enthusiasm for “Bayes’ method of determining probabilities a priori

by an experiment" (the title of his (1962)) Fisher could not pretend that the required "auxiliary experiment" for establishing the prior (the rolling of the W ball or something similar) would often be available.

References

- Aldrich, J. (1997). "R. A. Fisher and the Making of Maximum Likelihood 1912-22." *Statistical Science*, 12: 162-176. 161, 162
- (2000). "Fisher's "Inverse Probability" of 1930." *International Statistical Review*, 68: 155-172. 164
- (2005). "The Statistical Education of Harold Jeffreys." *International Statistical Review*, 73: 289-308. 163, 165
- (2008). "Keynes among the Statisticians." To appear in *History of Political Economy*. 163
- Barnard, G. A. (1987). "R. A. Fisher—A True Bayesian." *International Statistical Review*, 55: 183-189. 167
- Bayes, T. (1763). "An Essay towards Solving a Problem in the Doctrine of Chances." *Philosophical Transactions of the Royal Society*, 53: 370-418. 162, 166
- Bennett, J. H. (ed.) (1990). *Statistical Inference and Analysis: Selected Correspondence of R. A. Fisher*. Oxford: Oxford University Press. 166, 167
- Dale, A. I. (1990). *A History of Inverse Probability from Thomas Bayes to Karl Pearson*. New York: Springer. 164
- (2003). *Most Honorable Remembrance: The Life and Work of Thomas Bayes*. New York: Springer. 161
- Edwards, A. W. F. (1978). "Commentary on the Arguments of Thomas Bayes." *Scandinavian Journal of Statistics*, 5: 116-118. 163
- (1997). "What did Fisher Mean by "Inverse Probability" in 1912-1922?" *Statistical Science*, 12: 177-184. 162
- (2004). "Comment on D. R. Bellhouse "The Reverend Thomas Bayes, FRS: A Biography to Celebrate the Tercentenary of His Birth"." *Statistical Science*, 19: 34-37. 161, 162, 164
- Fienberg, S. (2006). "When did Bayesian Inference become "Bayesian"?" *Bayesian Analysis*, 1: 1-40. 161, 164
- Fisher, R. A. (1912). "On an Absolute Criterion for Fitting Frequency Curves." *Messenger of Mathematics*, 41: 155-160. 162

- (1921). “On the ‘Probable Error’ of a Coefficient of Correlation Deduced from a Small Sample.” *Metron*, 1: 3–32. 162, 164
- (1922). “On the Mathematical Foundations of Theoretical Statistics.” *Philosophical Transactions of the Royal Society, A*, 222: 309–368. 162, 163, 164
- (1922/3). “Review of J. M. Keynes’s Treatise on Probability.” *Eugenics Review*, 14: 46–50. 163
- (1925a). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd, 1 edition. 163
- (1925b). “Theory of Statistical Estimation.” *Proceedings of the Cambridge Philosophical Society*, 22: 700–725. 164
- (1930). “Inverse Probability.” *Proceedings of the Cambridge Philosophical Society*, 26: 528–535. 164
- (1932). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd, 4 edition. 164
- (1935a). *The Design of Experiments*. Edinburgh: Oliver & Boyd. 165
- (1935b). “The Logic of Inductive Inference, (with discussion).” *Journal of the Royal Statistical Society*, 98: 39–82. 165, 166
- (1936). “Uncertain Inference.” *Proceedings of the American Academy of Arts and Science*, 71: 245–258. 166
- (1956). *Statistical Methods & Scientific Inference*. Edinburgh: Oliver & Boyd, 1 edition. 163, 165, 166
- (1958). “The Nature of Probability.” *Centennial Review*, 2: 261–274. 161
- (1959). *Statistical Methods & Scientific Inference*. Edinburgh: Oliver & Boyd, 2 edition. 167
- (1960). “On Some Extensions of Bayesian Inference Proposed by Mr Lindley.” *Journal of the Royal Statistical Society, B*, 22: 299–301. 167
- (1962). “Some Examples of Bayes’ Method of the Experimental Determination of Probabilities A Priori.” *Journal of the Royal Statistical Society, B*, 24: 118–124. 163
- Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley. 161
- Howie, D. (2002). *Interpreting Probability: Controversies and Developments in the Early Twentieth Century*. New York: Cambridge University Press. 163, 165
- Jeffreys, H. (1935). “Contribution to the Discussion of Fisher (1935a).” *Journal of the Royal Statistical Society*, 98: 70–72. 166

- Laplace, P.-S. (1995). *Pierre-Simon Laplace: Philosophical Essay on Probabilities*. New York: Springer. 165
- Lindley, D. V. (1957). "Review of Statistical Methods and Scientific Inference by Sir Ronald A. Fisher." *Heredity*, 11: 280–283. 167
- (1958). "Fiducial Distributions and Bayes' Theorem." *Journal of the Royal Statistical Society, B*, 20: 118–124. 167
- Pearson, K. (1920). "The Fundamental Problem of Practical Statistics." *Biometrika*, 13: 1–16. 163
- Stigler, S. M. (1982). "Thomas Bayes's Bayesian Inference." *Journal of the Royal Statistical Society, A*, 145: 250–258. 163
- (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge MA: Harvard University Press. 161, 164
- (2005). "Fisher in 1921." *Statistical Science*, 20: 32–49. 161, 162
- Todhunter, I. (1865). *A History of the Mathematical Theory of Probability: from the Time of Pascal to that of Laplace*. London: Macmillan. 164
- Zabell, S. (1989). "R. A. Fisher on the History of Inverse Probability (with discussion)." *Statistical Science*, 4: 247–263. 161, 163

Acknowledgments

I am grateful to the editor and the referee for helpful suggestions in improving this paper.