

What Did Fisher Mean by “Inverse Probability” in 1912 – 1922?

A. W. F. Edwards

Abstract. The method of maximum likelihood was introduced by R. A. Fisher in 1912, but not until 1922 under that name. This paper seeks to elucidate what Fisher understood by the phrase “inverse probability,” which he used in various ways before defining “likelihood” in 1921 to clarify his meaning.

Key words and phrases: R. A. Fisher, inverse probability, likelihood, history of statistics.

In his 1922 paper “On the mathematical foundations of theoretical statistics” Fisher made a rather puzzling remark:

I must indeed plead guilty in my original statement of the Method of Maximum Likelihood (1912) to having based my argument upon the principle of inverse probability; in the same paper, it is true, I emphasised the fact that such inverse probabilities were relative only.

The remark is puzzling because in his 1912 paper Fisher is clear that the entity he is maximizing (not yet called the *likelihood*) “is a relative probability only, suitable to compare point with point, but incapable of being interpreted as a probability distribution over a region, or of giving any estimate of absolute probability.” Moreover, contrary to his assertion in 1922, the 1912 paper does not contain any argument as such, but merely the magisterial statement (after dismissing least squares and the method of moments), “But we may solve the real problem directly,” followed a few lines later by the assertion that “The most probable set of values for the [parameters] will make [the likelihood] a maximum.” (In the original the corresponding mathematical symbols were employed rather than the words in square brackets, but it will sometimes be convenient in this account to use Fisher’s later terminology anachronistically.)

G. A. Barnard has suggested to me that Fisher’s comments about relative probability in 1912 might

have been something of an afterthought; they are indeed confined to the sixth and last section of the paper. By contrast, the phrase “inverse probability system” is used in Section 5 to describe the graph of the likelihood function for the mean m and dispersion parameter h of a Normal distribution ($h = 1/[\sqrt{2} \sigma]$ in modern notation). Fisher says that Mr. T. L. Bennett, in a printed technical lecture, has integrated out m in order to derive a function of h to maximize for variation in h alone. However, “We shall see (in Section 6) that the integration with respect to m is illegitimate and has no definite meaning with respect to inverse probability,” a comment which might have been added at the time he drafted the additional Section 6. The interpretation is an attractive one, the more so because it explains the wording of the last sentence of all: “In conclusion I should like to acknowledge the great kindness of Mr. J. F. M. Stratton [*sic*; F. J. M. Stratton], to whose criticism and encouragement *the present form* of this note is due” (my italics).

A possible conclusion from Fisher’s two uses of the phrase “inverse probability” in 1912 is that he meant by it what he later called the likelihood, because (1) it was analytically equal to the likelihood and (2) it could not be integrated. Moreover the *principle* of inverse probability is not mentioned.

I only know of two published comments on the 1922 remark quoted above. Twenty years ago (Edwards, 1974a) I interpreted it as an admission by Fisher that he “was using the phrase “inverse probability incorrectly,” while in his recent introduction to the 1922 paper Geisser (1992) says that in 1912 “[Fisher] had taken a Bayesian approach because the maximising procedure resembled the calculation of the mode of a posterior probability.” I believe both of these comments are wide of the mark, mine

A. W. F. Edwards is Reader in Biometry, University of Cambridge, Department of Community Medicine, Forvie, Robinson Way, Cambridge CB2 2SR, United Kingdom (e-mail: awfe@medschl.cam.ac.uk).

because it does not reflect what Fisher actually wrote, and Geisser's because it does not mention Fisher's clear statement in 1912 that the probabilistic entity he was maximizing was not an ordinary probability but a kind of "relative" one which did not obey the addition law. The problem evidently needs looking at afresh, and in the present paper I shall try to examine exactly what Fisher meant by "inverse probability" in his youth. I should mention in passing that Zabell (1989), in his very informative paper "R. A. Fisher on the history of inverse probability," notes the 1922 remark, but simply takes it at its face value.

There is a famous comment of Fisher's from 1936 about "the theory of inverse probability," that "I may myself say that I learned it at school as an integral part of the subject, and for some years saw no reason to question its validity" (Fisher, 1936). Alas, we do not know how long "some years" were, though another historical remark two years later (published as a note to Jeffreys, 1938) gives a clue:

From a purely historical standpoint it is worth noting that the ideas and nomenclature for which I am responsible were developed only after I had inured myself to the absolute rejection of the postulate of Inverse Probability,

1. INVERSE PROBABILITY

Where does the phrase "inverse probability" come from, and has it always meant the same thing? It does not seem to have been used by Hume (1739), but Hume's contemporary David Hartley, in his *Observations on Man* (Hartley, 1749), wrote:

An ingenious Friend has communicated to me a Solution of the inverse Problem, in which he has shown what the Expectation is, when an Event has happened p times, and failed q times, that the original Ratio of the Causes for the Happening or Failing of an Event should deviate in any given Degree from that of p to q .

This, of course, is earlier than Bayes (1764), a fact which prompted Stigler (1983) to suggest that someone other than Bayes had discovered his theorem. The need for such an explanation only arises if the "Solution of the inverse Problem" is the Bayesian solution, and in response to Stigler I argued (Edwards, 1986) that it was not, writing "I myself doubt that the passage refers to the Bayesian

solution at all, believing it more likely to refer to one of the non-Bayesian attempts at a solution discussed first by James Bernoulli (1713) and then de Moivre (1738)." Dale (1988) agreed. When Todhunter came to write about *Ars Conjectandi* (Todhunter, 1865), he too wrote of "the *inverse* use of James Bernoulli's theorem" (his italics), whereas in his chapter on Bayes he did not use the word *inverse* at all (contrary to a statement of mine in Edwards, 1974a).

It is important to note that in this early use the word "inverse" refers to the problem itself, and not necessarily to a particular solution of it, and that when it does refer to a particular solution it might not be Bayes's. Richard Price, we may here remark, called the problem "the converse problem" in his introduction to Bayes's *Essay*.

Laplace (1774), in his "Mémoire sur la probabilité des causes par les évènements," gave what we now call Bayes's theorem (independently of Bayes) as his solution to the problem, but he seems never to have written of *inverse probability* (though Stigler, 1986, entitles his translation "Laplace's 1774 memoir on inverse probability"). Dale (1991) notes that the heading "Méthode inverse des probabilités" does occur in a Paris lecture summary from the turn of the century.

Augustus de Morgan, in the preface to his *Essay on Probabilities* (de Morgan, 1838), employed the phrase "the inverse method" to describe what is required if one is to reason "from the happening of an event to the probability of one or other cause." A little later in the preface he wrote:

De Moivre, nevertheless, did not discover the inverse method. This was first used by the Rev. T. Bayes, in *Phil.Trans.*liii.370.; and the author, though now almost forgotten, deserves the most honourable remembrance from all who treat the history of science.

Chapter III of the *Essay* is entitled simply "On inverse probabilities," and this is the earliest occurrence of the phrase of which I am aware. Boole (1854) in *The Laws of Thought* seems not to have used it, though Venn (1866) in *The Logic of Chance* did, quoting from de Morgan (1838).

We see from this brief summary that the interpretation to be placed on the word "inverse" changed with time, though the phrase "inverse probability," apparently introduced by de Morgan, carried (what we should now call) the Bayesian interpretation from the outset. It would, however, not be surpris-

ing to find some uncertainty as to what Fisher precisely meant by inverse probability in his early papers.

2. FISHER

We begin our exegesis of Fisher's post-1912 writings with the story of his controversy with Karl Pearson over the confusion between maximum likelihood and maximum posterior probability. This has been told too recently to need repeating in detail (E. S. Pearson, 1968; Edwards, 1974a), but we may scan it for clues about interpretation.

In his paper (Fisher, 1915) deriving the sampling distribution of the correlation coefficient Fisher also derived the maximum-likelihood estimate of the parameter. "I have given elsewhere [Fisher, 1912] a criterion, independent of scaling, suitable for obtaining the relation between an observed correlation of a sample and the most probable value of the correlation of the whole population." After the derivation he added "It is now apparent that the most likely value of the correlation will in general be less than that observed. . . ." Here "most probable" and "most likely" are evidently synonyms, but the word "inverse" is not used, and the criterion is "independent of scaling." Pearson and his collaborators were not clear about the distinction between maximum probability and Fisher's criterion, and their work (Soper et al., 1917) prompted Fisher to clarify his criterion by giving the word "likelihood" its technical meaning in 1921. Before then, however, there was another important interchange with Pearson.

The two letters, from the summer of 1916, are preserved in the Fisher archive in the Barr-Smith Library of the University of Adelaide and have been published by E. S. Pearson (1968). In the first, Fisher offers Karl Pearson the draft of a note for *Biometrika* commenting unfavorably on the method of minimum chi-squared which had been advocated by Kirstine Smith, a Danish pupil of Thiele's then studying under Karl Pearson (E. S. Pearson, 1990). Fisher's note ends:

There is nothing at all "arbitrary" in the use of the method of moments for the Normal curve; as I have shown elsewhere it flows directly from the absolute criterion ($\Sigma \log f$ a maximum) derived from the Principle of Inverse Probability. There is, on the other hand, something exceedingly arbitrary in a criterion which depends entirely upon the manner in which the data happens to be grouped.

Thus in mid-1916 Fisher has already formulated essentially the same perplexing statement as in 1922 about his criterion having been derived from the Principle of Inverse Probability.

The second letter is Pearson's reply, in which not surprisingly in view of Fisher's statement he does not differentiate between Fisher's criterion and maximum probability, which he now refers to as the Gaussian method. "If you will write me a defence of the Gaussian method, I will certainly consider it for publication, but if I were to publish your note, it would have to be followed by another note saying that it missed the point. . . ."

In 1918 Fisher submitted what was presumably his "defence," but after an interval Pearson rejected it in a letter dated 21 October 1918 (E. S. Pearson, 1968). Unfortunately no copy of Fisher's paper seems to exist, but probably it had something in common with the last section of Fisher (1921) and Section 12 of Fisher (1922).

It is to these two papers we must turn in order to find Fisher's clear detachment of the method of maximum likelihood from maximum posterior probability. The first of them had been rejected by Karl Pearson (his letter of 21 August 1920 is in E. S. Pearson, 1968) because he felt that "Under present printing and financial conditions, I am regretfully compelled to exclude all that I think erroneous on my own judgment, because I cannot afford controversy." Fisher was never to submit a paper to *Biometrika* again. Major Leonard Darwin, Fisher's mentor at this time, approached the Royal Statistical Society on his behalf to see if their *Journal* might be interested, but they could not help "because," so Dr. M. Greenwood informed him, "they have to cater for an audience many of whom could not understand it and they therefore have to limit the number of highly technical articles" (Box, 1978). However, in Italy Corrado Gini was looking for material for his new journal *Metron*, and it was there that the paper finally appeared. Thus the original definition of *likelihood* is in *Metron* and not *Biometrika* or the *Journal of the Royal Statistical Society*, the two leading British journals of the day.

In the Introduction Fisher explains how Soper et al. had incorrectly assumed that his criterion for estimation had been deduced from Bayes's theorem, and that in his opinion "two radically distinct concepts have been confused under the name of 'probability' and only by sharply distinguishing these can we state accurately what information a sample does give us respecting the population from which it is drawn." In Section 3 Fisher criticizes Soper et al. for having assumed that he had appealed to

Bayes's theorem in 1915, and adds:

As a matter of fact, as I pointed out in 1912 (Fisher, 1912) the optimum is obtained by a criterion which is absolutely independent of any assumption respecting the *a priori* probability of any particular value. It is therefore the correct value to use when we wish for the best value *for the given data*, unbiased by any *a priori* presuppositions.

The paper is dated October 1920. Within nine months, on 25 June 1921, the Royal Society received the manuscript of the 1922 paper with its statement "I must indeed plead guilty in my original statement of the Method of Maximum Likelihood (1912) to having based my argument upon the principle of inverse probability." The only way to reconcile these contemporaneous views of Fisher's about his own undergraduate paper nine years earlier is to suppose that he saw some distinction between the assumption of a uniform prior distribution and the principle of inverse probability, in accordance with the earlier, and looser, meaning attached to inverse probability which I discussed in the last section.

The 1921 paper ends with the "Note on the confusion between Bayes' Rule and my method of the evaluation of the optimum," in which *likelihood* is formally defined and differentiated from probability. (It was this heading which emboldened me to introduce the words *evaluate* and *evaluation* in 1972; I still think they ought to be adopted.)

Again in the 1922 paper Fisher repeatedly stresses the difference between likelihood, as newly defined, and probability. The phrase "method of maximum likelihood" occurs for the first time. He shows that he has read Bayes's paper with care, though he did not notice that Bayes had a cunning argument for adopting a uniform distribution for the binomial parameter (Molina, 1931; Edwards, 1974b, 1978). But he does not refer to Bayes's procedure as an example of the application of "inverse probability"; on the contrary, he says "In a less obtrusive form the same species of arbitrary assumption underlies the method known as that of inverse probability"; which he expounds for the case of two hypotheses. He states that the method assumes that their postdata probabilities are in the same ratio as the ratio of the probabilities of the data on the two hypotheses, and he notes that this amounts to assuming that the two hypotheses have been drawn at random from an infinite population in which each was true half the time. Then comes

his admission that in 1912 he had based his method of maximum likelihood on the principle of inverse probability. (The referees of the 1922 paper were A. S. Eddington, Plumian Professor of Astronomy, and G. Udney Yule, University Lecturer in Statistics, both in the University of Cambridge. Their reports are reproduced in Appendix 1.)

There is not much further evidence about the period 1912–1922 to be gleaned from Fisher's subsequent writings, but in the Adelaide archives there is a manuscript precis and discussion of Karl Pearson's paper "On the systematic fitting of curves to observations and measurements" (Pearson, 1902). The handwriting is that of Mrs. Fisher, so presumably she was taking dictation; there are two corrections in Fisher's hand. I reproduce the complete note in Appendix 2. The second paragraph reads:

It is noteworthy here, too, that throughout the paper no distinction is drawn between the fitting of frequency curves and that of regression lines. Only the latter had been traditionally treated by least squares. For the former the student might find in Gauss a discussion justifying what is now known as the method of maximum likelihood as a general principle, and showing that in fitting a normal frequency curve, this took the form of the method of moments, while in fitting regression lines in the important case of normal and equal variability in the arrays (i.e. of the observations) it took the form of the method of least squares. Gauss's views, though very influential, had not, however, in this matter gained general assent for he derived the method of maximum likelihood, erroneously, from the principle of inverse probability, confidence in which among mathematicians had been dwindling throughout the nineteenth century.

The word "erroneously" is perhaps not so placed as to convey quite the meaning Fisher intended, but the extract serves to confirm his knowledge of the work of Gauss and Pearson. Unfortunately it is undated, but it seems to record a study made for the 1922 paper though after the naming of likelihood, which would put it into the first half of 1921. This note does not differentiate between the principle of inverse probability and the adoption of a uniform prior distribution, which is how Gauss actually argued, yet Fisher made the distinction in the 1922 paper as I have indicated.

When Fisher (1930) introduced his notion of fiducial probability he called the paper “Inverse probability,” and it is indeed mostly a criticism of the Bayesian position along lines by now quite familiar, though he does add the new point that if we assume a uniform prior for a parameter, to choose its value by maximizing the posterior probability (as he again notes Gauss did) is very odd, for “had the inverse probability distribution any objective reality at all we should certainly, at least for a single parameter, have preferred to take the mean or the median value.” Laplace (1774) had called this mean “the mean of probability.” He also remarks, on introducing a fiducial distribution for the unknown parameter, “This is not inverse probability strictly speaking, but a perfectly direct argument, . . . ,” which suggests that he did not then reserve the phrase exclusively for Bayesian arguments.

Finally, we may note Fisher’s (1932) paper “Inverse probability and the use of likelihood” in which he was at pains to educate J. B. S. Haldane about the distinction between maximum probability and maximum likelihood, a task which he never fully accomplished (see Edwards, 1974a, 1996). He does not refer to his 1912 paper, but remarks acerbically that “Mathematicians have, however, often been tempted to apply the [Bayesian] procedure . . . to types of problem in which our *a priori* knowledge is certainly not of the definite kind postulated.” Throughout his life, Fisher used the word “mathematician” as code for someone who only understood deductive arguments, not inductive ones, and it is doubtful whether he included himself and his 1912 paper on this occasion.

3. CONCLUSION

We should never look for complete consistency in the writings of any author, especially a young one advancing the frontiers of a subject at a furious rate in the face of uncomprehending elders, but my impression now is that in the decade 1912–1922 Fisher did indeed draw a distinction between inverse probability and fully blown Bayesian inference which, though “of the same species,” starts from a slightly different viewpoint. Bayesian inference delivers a probability distribution for an unknown parameter, which Fisher explicitly and forcefully rejected from the start, while the principle of inverse probability only allows (on the present interpretation of his view) the comparison of parameter values “point with point” (Fisher, 1912). After all, if we cut away the historical use of the phrase, “inverse probability” is rather a good term

for “likelihood,” so long as we understand that it is not an ordinary probability. C. A. B. Smith (1986) observed that “Fisher’s choice of the word ‘likelihood’ might have been a little unfortunate, in that in ordinary language the words ‘likelihood’ and ‘probability’ are virtually synonymous.” I am inclined to agree with him; some kind of connotation of “support” might have been better (see Edwards, 1972).

APPENDIX 1

The following “General Remarks” of the referees of R. A. Fisher’s 1922 paper “On the mathematical foundations of theoretical statistics” are reproduced by kind permission of the President and Council of the Royal Society of London.

A. S. Eddington

The paper shows a remarkable insight into the theoretical ideas on which the methods of statistics ought to be based. I do not think that anyone else has arrived at so clear-sighted a view. The illustrative examples, and the application of the theoretical developments to a criticism of methods employed in practice are excellent; the criticism is fair, and shows the merits as well as the defects of previous work. It is an excellent paper, and I have no hesitation in recommending that it be presented as it stands.

G. Udny Yule

Mr. Fisher in this paper deals with the fundamental purpose of statistical measurement and the ‘efficiency’ of different methods of determining constants: he suggests a general method of determining “optimum” values [a method which has been used before but without, I think, specific recognition of it as a general method] and compares the ‘efficiency’ of other methods with this. A paper of such a basic kind is, I think, precisely the sort of paper which should be published by the Royal Society—speaking from the standpoint of the statistician. Someone else will, I hope, report from the standpoint of the mathematician.

APPENDIX 2

R. A. Fisher’s notes on Karl Pearson’s 1902 paper “On the systematic fitting of curves to observations and measurements” are reproduced by kind permission of the University of Adelaide.

This seems to be Pearson’s first attempt at a systematic paper on curve fitting; from its title one

would suppose that the problem to be discussed would be the estimation of unknown parameters, but actually its contents are much more miscellaneous. One opening remark requires comment.

p. 266. "So far I have not, however, been able to find any systematic treatise on curve-fitting. It is usually taken for granted that the right method for determining the constants is the method of least squares. But it is left to the unfortunate physicist or engineer to make the discovery that the equations for the constants found in this manner are in nine cases out of ten insoluble, or a solution so laborious that it cannot profitably be attempted."

The statement in the third sentence is an extraordinary one. Several hundred treatises on the method of least squares must have appeared between the writings of Gauss and the date of the quotation. Few if any of these can have failed to point out that by substituting an approximate solution the "normal equations" may be solved as simultaneous linear equations in the unknowns, and that, by repetition if necessary, the true solution may, apart from quite exceptional instances, be obtained with any required degree of precision. Evidently, Pearson had never had the experience of calculating a least square solution in any problem involving complicated functions of the unknowns. He must have formed an unfavorable opinion of the method without mastering any competent exposition of it, and have put forward the method of moments under a total misapprehension as to the difficulty or facility of applying the method accepted by his predecessors.

It is noteworthy here, too, that throughout the paper no distinction is drawn between the fitting of frequency curves and that of regression lines. Only the latter had been traditionally treated by least squares. For the former the student might find in Gauss a discussion justifying what is now known as the method of maximum likelihood as a general principle, and showing that in fitting a normal frequency curve, this took the form of the method of moments, while in fitting regression lines in the important case of normal and equal variability in the arrays (i.e., of the observations) it took the form of the method of least squares. Gauss's views, though very influential, had not, however, in this matter gained general assent for he derived the method of maximum likelihood, erroneously, from the principle of inverse probability, confidence in

which among mathematicians had been dwindling throughout the nineteenth century.

Later, p. 267, we read:

"I endeavour to show that it must give *good* values. The definition of "best fit" is more or less arbitrary, and for practical purposes, I have found that with due precautions as to quadrature, it gives, when one can make a comparison, sensibly as good results as the method of least squares."

Without a criterion of the best fit it is of course impossible to demonstrate that one method is good as or better than another.

The first section is a "general theorem" reading as follows: "A series of measurements or observations of a variable y having been made, corresponding to a series of values of a second variable x , it is required to determine a good method of fitting a theoretical or empirical curve $y = \varphi(x, c_1, c_2, c_3, \dots, c_n)$, where $c_1, c_2, c_3, \dots, c_n$ are arbitrary constants, to the observations for a given range $2l$ of the variable x ."

This "proof" occupies five pages and is heavily algebraic. In essence it is as follows. If the function φ is a polynomial of degree $n - 1$ the method of moments is the same as the method of least squares "This obviously gives a very good method, if not 'the best,' a term incapable of definition." φ is assumed to be expandible in a MacLaurin series within the required range. The remainder after the term in x^{-1} is said by hypothesis to be small. In consequence its differential coefficients with respect to the unknowns are neglected. The author shows no consciousness that to say a quantity is small means little unless he can say that it is small compared with some other quantity with which it is relevant to compare it.

The second section give some of the rules for mechanical quadrature. The third gives the results of fitting a curve of the third degree to the interval between the first and last of eleven equidistant observations. It is not stated that to fit by least squares in such a case is much easier than to find the areal moments by quadrature. Of the four fittings by moments given three do badly, indeed two of them give more than double the residual variance of least squares but that using Sheppard's rule is only about 7% larger.

In Section 4 we come to fitting frequency curves but it is devoted principally to the development of quadrature formulae to allow for grouping. The remaining sections give illustrations of (i) fitting a Pearsonian type 1 curve to the fecundity of broodmares, (ii) fitting a Pearsonian type 3 curve to a *discontinuous* frequency distribution given by Thiele for a game of patience, (iii) fitting sine curve of arbitrary period to the data used above for the curve of the third degree and (iv) fitting Makeham's formula to mortality data from 25 to 85 years.

ACKNOWLEDGMENTS

This investigation was stimulated by a visit to the University of Adelaide to study the papers of R. A. Fisher. I am grateful to the Royal Society of London for a study grant for the visit, and to the Institute of Mathematical Statistics of the University of Copenhagen for the invitation to participate in the symposium in honor of Professor Anders Hald in June 1993. The present paper is a reduced and corrected version of the preprint issued by the Institute in 1994, "Presented to Professor Anders Hald in Honour of His Eightieth Birthday" (Universitetsparken 5, DK-2100, Copenhagen Ø, Denmark), and owes much to the critical comments of J. C. Aldrich, G. A. Barnard, A. I. Dale, A. Hald, G. Shafer and S. ZABELL, to all of whom I am most grateful. Responsibility for the interpretations is, however, mine alone.

REFERENCES

- BAYES, T. (1764). An essay towards solving a problem in the doctrine of chances. *Philos. Trans. Roy. Soc.* **53** 370–418. [Reprinted in *Studies in the history of probability and statistics IX: Thomas Bayes' essay towards solving a problem in the doctrine of chances* (with a biographical note by G. A. Barnard) *Biometrika* **45** 293–315 (1958) and in Pearson and Kendall (1970).]
- BERNOULLI, J. (1713). *Ars Conjectandi*. Thurnisius, Basilea. [Facsimile reprint, *Culture et Civilisation*, Bruxelles (1968). Reprinted in *Die Werke von Jakob Bernoulli* **3** 107–286, Birkhäuser, Basel (1975).]
- BOOLE, G. (1854). *An Investigation of the Laws of Thought*. Walton and Maberley, London. [Reprinted (1958) by Dover, New York.]
- BOX, J. F. (1978). *R. A. Fisher: The Life of a Scientist*. Wiley, New York.
- DALE, A. I. (1988). On Bayes' theorem and the inverse Bernoulli theorem. *Historia Math.* **15** 348–360.
- DALE, A. I. (1991). *A History of Inverse Probability*. Springer, New York.
- DE MOIVRE, A. (1738). *The Doctrine of Chances*, 2nd ed. Woodfall, London. [Reprinted (1967) by Cass, London.]
- DE MORGAN, A. (1838). *An Essay on Probabilities and Their Application to Life Contingencies and Insurance Offices*. Longman, Orme, Brown, Green, & Longmans, London.
- EDWARDS, A. W. F. (1972). *Likelihood*. Cambridge Univ. Press. [Reprinted in Edwards (1992).]
- EDWARDS, A. W. F. (1974a). The history of likelihood. *International Statistical Review* **42** 9–15. [Reprinted in Edwards (1992).]
- EDWARDS, A. W. F. (1974b). A problem in the doctrine of chances. In *Proceedings of the Conference on Foundational Questions in Statistical Inference* (O. Barndorff-Nielsen, P. Blæsild and G. Schou, eds.) 43–60. Aarhus Univ. [Reprinted in Edwards (1992).]
- EDWARDS, A. W. F. (1978). Commentary on the arguments of Thomas Bayes. *Scand. J. Statist.* **5** 116–118.
- EDWARDS, A. W. F. (1986). Is the reference in Hartley (1745) to Bayesian inference? *Amer. Statist.* **40** 109–110.
- EDWARDS, A. W. F. (1992). *Likelihood* (expanded edition). Johns Hopkins Univ. Press.
- EDWARDS, A. W. F. (1996). The early history of the statistical estimation of linkage. *Annals of Human Genetics* **60** 237–249.
- FISHER, R. A. (1912). On an absolute criterion for fitting frequency curves. *Messenger of Mathematics* **41** 155–160.
- FISHER, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **9** 507–521.
- FISHER, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron* **1** 3–32.
- FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **222** 309–368.
- FISHER, R. A. (1930). Inverse probability. *Proc. Cambridge Philos. Soc.* **26** 528–535.
- FISHER, R. A. (1932). Inverse probability and the use of likelihood. *Proc. Cambridge Philos. Soc.* **28** 257–261.
- FISHER, R. A. (1936). Uncertain inference. *Proceedings of the American Academy of Arts and Science* **71** 245–258.
- GEISSER, S. (1992). Introduction to Fisher (1922) On the mathematical foundations of theoretical statistics. In *Breakthroughs in Statistics 1. Foundations and Basic Theory* (S. Kotz and N. L. Johnson, eds.) 1–10. Springer, New York.
- HARTLEY, D. (1745). *Observations on Man, His Frame, His Duty, and His Expectations*. Richardson, London. [Reprinted (1966) by Scholar's Facsimiles and Reprints, Gainesville, FL.]
- HUME, D. (1739). *A Treatise of Human Nature* (L. A. Selby-Bigge, ed., 1888 and later printings). Clarendon, Oxford.
- JEFFREYS, H. (1938). Maximum likelihood, inverse probability, and the method of moments. *Annals of Eugenics* **8** 146–151.
- LAPLACE, P. S. DE (1774). Mémoire sur la probabilité des causes par les évènements. [Translated in Stigler (1986).]
- MOLINA, E. C. (1931). Bayes' theorem: an expository presentation. *Ann. Math. Statist.* **2** 23–37.
- PEARSON, E. S. (1968). Some early correspondence between W. S. Gosset, R. A. Fisher and Karl Pearson, with notes and comments. *Biometrika* **55** 445–457. [Reprinted in Pearson and Kendall (1970).]
- PEARSON, E. S. (1990). 'Student': A Statistical Biography of William Sealy Gosset (R. L. Plackett, ed., with the assistance of G. A. Barnard). Clarendon, Oxford.
- PEARSON, E. S. and KENDALL, M. G. (1970). *Studies in the History of Statistics and Probability*. Griffin, London.

- PEARSON, K. (1902). On the systematic fitting of curves to observations and measurements. *Biometrika* **1** 265–303.
- SMITH, C. A. B. (1986). The development of human linkage analysis. *Annals of Human Genetics* **50** 293–311.
- SOPER, H. E., YOUNG, A. W., CAVE, B. M., LEE, A. and PEARSON, K. (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.
- STIGLER, S. M. (1983). Who discovered Bayes's Theorem? *Amer. Statist.* **37** 290–296.
- STIGLER, S. M. (1986). Laplace's 1774 memoir on inverse probability. *Statist. Sci.* **1** 359–378.
- TODHUNTER, I. (1865). *A History of the Mathematical Theory of Probability*. Macmillan, Cambridge and London. [Reprinted (1965) by Chelsea, New York.]
- VENN, J. (1866). *The Logic of Chance*. Macmillan, Cambridge and London.
- ZABELL, S. (1989). R. A. Fisher on the history of inverse probability. *Statist. Sci.* **4** 247–263.