Bayes, based on, and appropriate to, expectations in games of chance, and in the case of other concepts distinct from this, yet which may seem sufficiently analogous to cause confusion or which are equally relevant in processes of inductive inference, I have preferred to seek for distinct appropriate names, of which 'likelihood' and 'quantity of information' may serve as examples.

The passage of Laplace to which you kindly draw my attention is one which I had occasion to quote some years ago in addressing the Tercentenary Conference at Harvard [*CP* 137]. I think it does show that Laplace was not prepared to appreciate the caution and scepticism which actually prevented Bayes from publishing his treatise during his lifetime. It was, as you know, published posthumously at the instance of his friends.

H. Gray to Fisher: 15 *June* 1951

The bearer of 'news', and the candid friend, make themselves unliked; but I am willing to take the risk. The *Scientific Monthly* may not come under your notice, being aimed at laymen, *Science* at professionals; both publ. by Am. Assn. Advancement of Science. Also you may not care to notice Neyman's notions.[1] If it annoys, forgive me.

[1] Neyman's review of *CMS* in *Sci. Month.* **72**, 406–8, (1951).

Fisher to H. Gray: 2 *July* 1951

The second word of the phrase 'candid friend' is sometimes not taken very seriously. If I mistake not, however, you mean it to be in this case. Hence what follows:—

Neyman is, judging by my own experience, a malicious mischief-maker. Probably by now this is sufficiently realised in California. I would not suggest to anyone to engage in scientific controversy with him, for I think that scientific discussion is only profitable when good faith can be assumed in the common aim of getting at the truth.

There are, however, in the review two points in which he seeks to impugn my own good faith, and in such a case it may have seemed to you a friendly office to show that some of my friends in California do not share his opinion.

(a) The use of Euclidean hyperspace of N dimensions to represent a random sample of any size N. Before ascribing this innovation to me, Mahalanobis[1] had looked into the matter. His claim, of course, is not that I invented Euclidean hyperspace, but that this particular use of it enabled me, starting in 1915, to give the exact solutions of a number of problems of distribution, which had been unsuccessfully sought by other means. In 1900 Pearson had used hyperspace to represent the correlational properties of a number of different variates. Each variate was given one dimension. This is not the same as discussing a finite sample and assigning a dimension to each observation. If the latter process had occurred to Pearson fifteen years earlier, it would be curious that neither he, nor the abler mathematicians (e.g. Sheppard and Soper) associated with him at the time, had found the solutions they were seeking; which indeed are nearly obvious, and come out at once, when my method is used.

I can see no source for Neyman's imputation save a desire to be spiteful. His whole exposition of tests of significance is based on my 'sample space', but such indebtedness he is not inclined to admit. Likewise his 'power function' is, in disguise, a special application of my Likelihood.

(b) (see *J.R.S.S. 98*, p. 77)[2] Edgeworth's paper[3] of 1908 has, of course, been long familiar to me, and to other English statisticians. No one could now read it without realising that the author was profoundly confused. I should say, for my own part, that he certainly had an inkling of what I later demonstrated. The view that, in any proper sense, he anticipated me is made difficult by a number of verifiable facts. (i) He based his argument on Bayesian inverse probability; my results are free from this assumption and represent an entirely different approach. (ii) He ends his paper by explaining that he has been writing only of measures of central tendency, and not of 'the fluctuation', nor, it may be presumed, of measures of correlation, etc. (iii) The formula common to his work and mine, that for the variance of an efficient statistic, he obtained from Pearson and Filon (1898),[4] who believed it to give the sampling variance of statistics obtained by the method of moments. Already in 1903, using a direct method due to Sheppard, these erroneous values had been corrected, but Edgeworth does not refer to this, and may not have known of Sheppard's work. Obviously Sheppard's calculations had raised the questions:— Had Pearson and Filon's variances any validity at all? Does any class of estimate actually have these variances? If so, how can such an estimate be obtained in general? But Edgeworth would have been far ahead of his time had he asked them.

Setting aside the disputed postulate of inverse probability, and having regard to parameters of all kinds, irrespective of what quality of the population they measured, I showed in 1922 that statistics could in general be found with variances given by this formula, and that it was the least limiting variance possible.

Anyone wishing to ascribe my results to Edgeworth should at least have ascertained that he accepted them, which so far as I know he never did.

[1] Mahalanobis, P.C. (1938). Professor Ronald Aylmer Fisher. *Sankhyā* **4**, 265–72.
[2] Fisher's reply to discussion of his 1935 paper, 'The logic of inductive inference' (*CP* 124). See *Collected papers of R.A. Fisher*, Volume III, p. 310.
[3] Edgeworth, F.Y. (1908). On the probable errors of frequency constants. *J. R. Statist. Soc.* **71**, 651–78.
[4] Pearson, K. and Filon, L.N.G. (1898). Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Phil. Trans.* A **191**, 229–311.

Fisher to R.F. Harrod: 3 *December* 1956

My new book this year, *Statistical Methods and Scientific Inference*, coincides in time so closely with yours[1] that I was more than a little anxious to ascertain what point of view it was that you were expressing, and as I think I indicated, I found from such soundings as I have made that I was, allowing for differences in the problems being discussed and occasionally in the meanings attached to the words used, very much indeed in agreement with your point of view. When you care to glance at mine I hope you will be amused at the indications which I try to give, from several different points of view, of the reasons why mathematicians have cut such a comparatively poor figure in this business.

The two books seem to me to be frankly complementary. You are developing the subject in the philosophic tradition, and I almost exclusively in the mathematical; namely, I am concerned with the exact quantitative treatment of particular cases in which uncertain inference seems to be possible, with the specification for such cases for both the nature and the extent of the uncertainty. The book is sprinkled with warnings against hasty generalizations.

Quite naturally we use the word 'probability' rather differently, I with my eye fixed on those noble and ardent gamblers for whom Montmort and de Moivre were writing, and I am therefore concerned with, *senso strictu*, Mathematical Probability as developed by these mathematicians. Keynes's phrase 'measure of the degree of rational belief' obviously covers a much wider field which would shatter the old bottle of mathematical probability. It is, however, an admirable phrase, and I show that in many instances objective quantitative measures of this kind occur in situations where any statement of Mathematical Probability would be invalid.

Somewhere I noticed you turn aside to stick a pin into the word 'likelihood'. I should like to leave it to you to judge, after reading what you care to of my book, whether the use of this term is really unsound.

[1] Harrod, R.F. (1956). *Foundations of inductive logic*. Macmillan, London.

R.F. Harrod to Fisher: 8 *January* 1957

I have read through your book once. I curse myself for not having insisted on being taught more mathematics at school. It is difficult to pick these things up late in life. I cannot pretend that I have understood more than about half the book at most. But there are matters on which I hope I may still gain an understanding through better acquaintance with some of the forms of art. I am ordering *Statistical Methods*. Also I want to understand more about the 'fiducial argument' which is central to so much in your book. According to the refs. I think I ought to read your 1930 article.

I think I was in sympathy with the greater part of your thought—and much enjoyed your controversial passages. And I naturally rejoiced when I found, as I sometimes did, passages which almost echoed certain things in my book.

If you have read that, I fear you will be deeply shocked at my crudity and sloppiness whenever I approach the fringe of mathematical problems. My impulse, as you say, was a philosophical one. You may not realize how completely the philosophical world has given up induction in despair. Philosophers teach their pupils that the validity of inductive reasoning can in no wise be established; we are bidden to admire the *de facto* success of scientists and to take their methods on trust. They are to be justified solely by their pragmatic success. That would incidentally put your work out of court completely! You are under an illusion, they must suppose, if you think that by studying the logic inherent in certain methods, you can actually improve those methods. For according to them there is *no* logic inherent in them.

At Oxford the majority of clever boys not reading Maths or Natural Science do philosophy, whether in Greats (Classics leading on to modern philosophy) or in Modern Greats (Philosophy, Politics and Economics). All these are sent out into the world fully convinced that induction is a bit of fraudulent trickery, which just happens to work. I think this bad. It means that this large group, who have some influence on our affairs, are basically anti-rational.

Philosophy is in the same state everywhere. The only philosopher I can find who believes that there is any validity in induction is the man, Williams of Harvard, to whom I have referred. Unfortunately his book is not a very good one, containing much rather loose controversializing.

I believe that your fiducial argument, if only I could grasp its essence aright, will be found to require something like my fair sampling postulate. The philosophers are up in arms saying that we have no right to suppose with any degree of probability, however low, that our successive experiences do not consist of events which, given the real condition of nature, are excessively improbable. That is why I think my 'principle of experience' to be of real logical value, *viz.* the principle that one is not likely to be on the edge of a great surface. For this has the happy consequence that it does not matter whether the fair sampling postulate that we assume in making inductions is true or not.

I am sorry about my footnote on likelihood! What I had in mind was philosophical writing in which the distinction was between objective and subjective, or, if not subjective, then at least purely intuitional probability. If I have understood aright, your likelihood is fully as objective as probability.

The passage where I felt least happy was on pp. 41–2.[1] I wholly agree with your remarks on p. 110[2] about confusion in regard to infinity and the limit (and have said something similar). But I do want to define probability as arising from data of a certain *logical* character (e.g. a sample of a certain size and sort) that ensures that beliefs arising by inference from the data shall be more often right than wrong. This seems to imply a set of occasions on which one is confronted with data of that logical character. And I would *prefer* not to assent to your indifference (p. 109)[3] as between object, event and proposition as the predicand of probability statements; the objects or events you have in mind can presumably always be converted into contents of propositions. I do want to keep probability as a logical relation, *viz.*, between data and the inference from them. Incidentally I have studiously avoided any reference to 'propositions', always writing in terms of 'beliefs', so as to avoid being bogged down by the symbolic logicians and their tortured paradoxes. . . .

[1,2,3] Pages 44-5, 114, and 113 respectively, in *SMSI*.

Fisher to R.F. Harrod: 15 *January* 1957

It was nice of you to write me such a long letter. For my part I am quite as lost in philosophical terms as you can be in mathematical. In different languages we do seem to be doing very much the same thing, you being primarily concerned at a much deeper level, and in more qualitative terms, with what I am concerned with in trying to give an exact mathematical formulation of the steps from data to inference in cases where, *more mathematicorum*, I choose carefully what sorts of data I am prepared to discuss, and so am working at a more accessible level.

Both of us are building, I believe, fundamentally, on the logical force of what I call a test of significance, an inferential force which I describe as more elemental than that of a statement of probability; and here I am using 'probability' in a stricter, i.e. less inclusive, sense than Keynes' 'degree of rational belief', because I am concerned to distinguish the historical Mathematical Probability from various other measures of rational belief appropriate to circumstances in which no mathematical probability can be asserted. This seems to be a rather remarkable conformity of intention, scarcely at all frustrated by the differences attached to the meanings of the words used.

May I take now some of the interesting points you raise, not that any of them seem to involve a fundamental difference but because their clarification may well lead to a better understanding.

Thank you for your corrections; I am returning a few that I have found, including some improvements in the wording that have occurred to me. The only one of yours needing discussion is where you refer to the statement on page 19 [*SMSI*], ' The expectation from a mating between two heterozygotes is 1 homozygous black, to 2 heterozygotes, to 1 brown'. My intention in saying this is that the numbers enumerated should be taken out of their total of four, so that 1/4 are said to be homozygous black, 1/2 heterozygous black, and 1/4 brown. I expect the way I put it is elliptical and should be expanded. Its substance, however, is merely to lead to the justification of the prior probability ratio of 1 homozygote to 2 heterozygotes for a black mouse from such a mating. I agree entirely that probability, either in the narrow sense of the eighteenth century gamblers or in any wider philosophical sense you may prefer to use, refers to strength of rational belief properly to be inferred from given data. On page 41–42,[1] however, I am trying to say that we also have rational grounds for belief in many cases in which probability (or frequency) statements cannot properly be made, and that this is typical of simple tests of significance, from which, in my narrow sense, no probability statements can properly be inferred, although I do point to a somewhat peculiar class of cases in which, not by a simple test of significance but by a more sophisticated fiducial argument, true probability statements are available. They do not in my opinion make the inference much stronger, though they do make it somewhat stronger in respect to the specification of precision as elaborated in Section 4, as contrasted with, for example, that of the likelihood graphs on page 72.[2]

I agree with you and with Keynes that the predicand of a probability statement can always be taken to be a proposition, but I would plead that the popular use which applies the term to objects or events is also correct and sufficiently vivid to supply a warning of the kind that I think Keynes needed in the remark I quote on page 44.[3]

I do not think I have referred much to propositions, but I have often used the term 'assertions' where you use 'beliefs', because I am not concerned with beliefs in themselves and as such, but with their communication to other rational minds.

[1,2,3] Pages 44–5, 76, and 48, respectively, in *SMSI*.

W.E. Hick to Fisher: 6 *October* 1951

Several years ago I ventured to ask you some elementary questions about statistical inference, and your kindness then has encouraged me to approach you again, on a matter which, though elementary, I feel is of some practical importance.

The question has to do with the theory of significance tests. Most experimenters, of course, have neither the time nor the inclination to go deeply into the theory of critical regions, power functions, etc. Yet many (of those who think at all!) are puzzled about when to use 'one tail' or 'two tails' and why one should compute the probability of a sample 'as unlikely as, *or more unlikely than*. . .'. Involved in all this is our *bête noire*—chi-squared, with its deceptive simplicity.

The question the experimenter thinks he is asking himself is surely, 'What is the probability, on the chance hypothesis, that I should have got the result I did get?' It is difficult to tell him straightforwardly why he should care about the probabilities of all the other, less likely, results that he did *not* get. . . .

Fisher to W.E. Hick: 8 *October* 1951

I am a little sorry that you have been worrying yourself at all with that unnecessarily portentous approach to tests of significance represented by the Neyman and Pearson critical regions, etc. In fact, I and my pupils throughout the world would never think of using them. If I am asked to give an explicit reason for this I should say that they approach the problem entirely from the wrong end, i.e. not from the point of view of a research worker, with a basis of well grounded knowledge on which a very fluctuating population of conjectures and incoherent observations is continually under examination. In these circumstances the experimenter does know what observation it is that attracts his attention. What he needs is a confident answer to the question 'Ought I to take any notice of that?'. This question can, of course, and for refinement of thought should, be framed as 'Is this particular hypothesis overthrown, and if so at what level of significance, by this particular body of observations?'. It can be put in this form unequivocally only because the genuine experimenter already has the answers to all the questions that the followers of Neyman and Pearson attempt, I think vainly, to answer by merely mathematical consideration. In fact the practical experimenter does not often put up a damn-fool test of significance but it is a labour of many years and much art for the 'Theory of Testing Hypotheses' to avoid such tests.

Fisher to H. Hotelling: 3 *October* 1941

I have just read your letter of September 5th with pleasure, for it is always a pleasure to have a letter from you. I am very glad you noticed E.B. Wilson's article,[1] for he is a man of such generally sound judgment that one could not have expected that the test of significance he advocates for the four-fold table would have appealed to him. The point is, however, really rather a knotty one, for he might have adduced as the ground for choosing $p = \frac{1}{2}$, not that it was the value observed for the total of experimental and controlled animals, but that this value maximises the aggregate probability of the six possible outcomes of the experiment in which the surviving controls exceed the surviving experimental mice by four or more. Such a criterion would, of course, work out differently in other cases, and would be extremely troublesome to apply, but it would, I think, also be extremely troublesome to argue against.

With respect to Behrens' test, I should agree with you that Behrens' own statement seems extremely confused and unconvincing. I should not rely on it without developing an independent argument leading to the same result. I should like to be clear, however, that, far from ignoring the connection between t and s in samples from populations having a given value of σ, I regard this connection as essential to the development either of Behrens' or of 'Student's' test. Given σ, in fact, the distribution of t is not independent of s, and what is wanted is the product ts, the distribution of which depends on σ only. It is only because the Gaussian test, using the probability integral of the normal curve, is not available for rigorous purposes when σ is not known, that the observable quantities t and s come into consideration.

The logical development of the rest of the argument depends, so far as I can see, on three points, two of which are mathematically demonstrable, while the third, which justifies their relevance, is a matter of judgment.

(a) In random samples of n' from normal populations the ratio s/σ has a distribution dependent only on n', and therefore known.

(b) All other information supplied by the data is irrelevant to the value of σ. This is the sufficiency property.

(c) (which is, I think, undoubtedly incapable of mathematical proof). The sample under examination is unbiassed in respect of the values $\bar{x}$ and s obtained, so that for the population of possibilities to which our probability statement shall refer, the ratio s/σ takes all its possible values with the theoretical frequencies, independently of $\bar{x}$, or the fact that we can use the mathematical short cut of calculating the distribution of $\bar{x}/s$, and using in probability statements probabilities derived from its distribution. This act of judgment in the logical basis of 'Student's' test has, I think, obscured the issue to mathematicians much more than to practical experimenters. I think, however, that an analogous act of judgment enters into all applications of probability theory.

Of course this is, as Behrens' test shows, incompatible with any theory of tests of significance, such as that of Neyman, which makes it axiomatic that a given percentage of the observations in repeated trials shall be found to be significant. It allows for the fact that some of the unknown parameters may have the effect of making the test more or less sensitive, so that on fixing one of these we should find that the percentage of cases recognised as significant differed from the percentage with which the test of significance was to be entered. As I understand it, however, a test of significance is a logical statement to the effect that either a well defined event having a rarity definable in terms of the observations has occurred, or else that a certain hypothesis is untrue. Events the rarity of which can only be defined in terms of unknown parameters cannot in this sense provide tests of significance.

I hope you will write further about all this.

[1] Wilson, E.B.(1941). The controlled experiment and the fourfold table. *Science* **93**, 557–60.

Fisher to J.O. Irwin: 8 *November* 1935

I am glad you have given me an opportunity of reconsidering the arguments in the neighbourhood of page 721 of my paper on Statistical Estimation [*CP* 42]. The main point, which I think only becomes clear when the paper is read as a whole, is that we are concerned with the difference between two systems of equi-statistical surfaces or regions, and usually not with the difference between different regions of the same system. I am returning your letter for reference, together with a few sheets which I hope clear up the main difficulty, though I don't discuss the analytic handling of the variance of the second degree deviations between corresponding surfaces of the different systems. I will go into this with pleasure if you would like me to.

[Enclosure]

1. The origin of the loss of information incurred by using an estimate from a sample, in conjunction with the sampling distribution of all such estimates from samples of a given size, has been traced to the fact that the equistatistical surfaces on which $\partial L/\partial\theta = 0$ for different values of θ do not coincide with the system of surfaces on which $\partial L/\partial\theta = C$ for any one value of θ. The systems clearly have one surface in common, and the neighbouring surfaces of the two systems will generally be inclined to each other at only small angles; e.g. if $\hat\theta - \theta$ is small, the surface of the first system $\partial L/\partial\theta = 0$ may be chosen, with a certain harmless arbitrariness, to correspond with one of the second system, such as that which meets it on the locus of expectation

$$\frac{\partial L}{\partial\theta} = S\left\{\frac{\hat m}{m}\ \frac{\partial m}{\partial\theta}\right\}.$$

Then my assertion is that the angle between these corresponding surfaces of the two systems will, in the neighbourhood where $\hat\theta$ is near to θ, be proportional to the difference $\hat\theta - \theta$. The equations of the two surfaces are in fact

$$S\left\{\frac{x}{\hat m}\ \frac{\partial\hat m}{\partial\theta}\right\} = 0,$$

and

$$S\left\{\frac{x}{m}\ \frac{\partial m}{\partial\theta}\right\} = S\left\{\frac{\hat m}{m}\ \frac{\partial m}{\partial\theta}\right\},$$

and the coefficient of x in the first may, if the differential coefficients of log m are finite, be expanded as

$$\frac{1}{m}\ \frac{\partial m}{\partial\theta} + (\hat\theta - \theta)\frac{\partial^2}{\partial\theta^2}\log m + \frac{(\hat\theta-\theta)^2}{2}\ \frac{\partial^3}{\partial\theta^3}\log m + \ldots$$

Using the convention that the square of the cosine of the angle between surfaces

$$S(px) = C,$$

and

$$S(p'x) = C',$$

is

$$\cos^2\alpha = \frac{S^2(mpp')}{S(mp^2)S(mp'^2)}$$

one can obtain

$$\sin^2\alpha = \frac{(\hat\theta-\theta)^2}{\mathrm{I}} S\left\{\frac{1}{m}\left(m'' - \frac{m'^2}{m}\right)^2\right\} + \text{smaller terms}.$$

2. It being agreed that we have to evaluate the variance of $\partial^2L/\partial\theta^2$ over regions for which $\partial L/\partial\hat\theta = 0$, the question is whether it is sufficient to evaluate it within the regions for which $\partial L/\partial\theta = C$. Might one not argue as follows? The variance within any one of the regions $\partial L/\partial\hat\theta = 0$ will be the same as that evaluated, together with a corrective term proportional to $\hat\theta - \theta$, and other terms of higher order. Averaging all these according to their frequency of occurrence in the limit for large samples, we obtain a corrective term proportional to the variance of $\hat\theta - \theta$, or of order $1/n$ compared with the value obtained.

I think this also covers the logical though not the analytical difficulty on your p. 3.

With respect to p. 4 one might, perhaps, put the position in this way:— By basing our notion of precision only on a single estimate and the size of the sample, we are not neglecting such information about precision as our estimate provides, but we are neglecting other information provided by the sample respecting the precision of an estimate of given magnitude derived from it. Common formulae for standard errors, for example, are of the form

$$\sigma = f(\theta, n).$$

Consequently, variations in precision associated with variations in θ are taken into account through the magnitude of our estimate, but variations in precision for samples of the same size and giving the same estimate are neglected.

Fisher to A.T. James: 4 *October* 1961

It was nice to get your letter which I found waiting for me when I got back here yesterday.

I shall soon be sending out two papers from *Sankhyā* which indeed you may already have seen, extending to other problems the method by which Behrens' test was justified in 1935. I think you will be interested in the derivation of the simultaneous distribution of the variances, when the means are known to be equal. I do not know how, otherwise, this distribution could have been obtained. Of course, all three problems are inaccessible to Neyman and Pearson's approach to testing hypotheses.

I have not done anything further with the problem of the simultaneous set of correlations among t normal variates. The simultaneous distribution is not difficult, and I ought to publish it, but I have not fished it out since coming back, and it may not be invertible, though I feel it ought to be. The example I take up first shows that Exhaustive Estimation is not strictly necessary, and it may not indeed be sufficient, though it would be surprising if it were not. The example with two variates shows that it is not too slick analytically.

I think consultant work is immensely important, but on condition that the consultant is concerned to learn, and not merely to pontificate.

A. T. James to Fisher: 14 *January* 1962

. . . Your discussion in 'Sampling the reference set' [*CP* 284] made it clear to me that in the test of the difference of means with unknown variance ratio, the distribution of

$$d = (\bar{x}_1 - \bar{x}_2) \Big/ \sqrt{\frac{s_1^2}{n_1+1} + \frac{s_2^2}{n_2+1}}$$

in the conditional distribution given s_1/s_2 averaged over the fiducial distributon of σ_1/σ_2 is the same as the distribution of

$$\left(\frac{s_1 t_1}{\sqrt{n_1+1}} - \frac{s_2 t_2}{\sqrt{n_2+1}}\right) \Big/ \sqrt{\frac{s_1^2}{n_1+1} + \frac{s_2^2}{n_2+1}}$$

. . .

Fisher to A. T. James: 13 *March* 1962

Your letter of Jan 14 has got around to me at last.

I am glad you liked the paper on Sampling the Reference Set. The business you call averaging out the distribution of σ_1/σ_2 was only an experimental way of putting the fact that the distribution of σ_1/σ_2 is exhaustively estimated by s_1/s_2 independently of $\bar{x}_1$, $\bar{x}_2$ and on this marginal distribution is superposed the array distributions of μ_1 and μ_2 for each σ_1/σ_2 that can be encountered. What I liked best and what rather surprised me is that the same experimental mechanism should give the simultaneous distribution of σ_1 and σ_2. I wish one could always set up the reference set so explicitly. . . .

I am not so sure that the tide of real feeling in the U.S. is so backward as it was 15 years ago, when so many leading posts fell to Neymanians. I think a good many are dissatisfied with mere 'decisions', which essentially evade the problem of specifying the nature of uncertainty, or the true grounds for belief. . . .

Fisher to G.S. James: 13 *August* 1955

Thank you for your letter.[1] It has been interesting to me to see the recent development of interest in Behrens' solution to this problem, due, I suppose, to recognition that the grounds on which it was at first attacked, principally by Bartlett under the influence of J. Neyman, are no longer tenable, for I suppose few except Egon Pearson now believe that the strength of the significance under which a composite hypothesis can be rejected is to be identified with the frequency with which such evidence is obtainable, or indeed that the various simple hypotheses comprising the composite will necessarily yield the same frequency. As a by-product, however, of this misunderstanding, a number of efforts have been made to discredit the mode of argument which I have termed 'fiducial'. I do not believe that practical men have ever had any difficulty in understanding, or doubts about using, Behrens' test, and during the current year I have been notified of new tabulations to facilitate its use. There is a good deal of material for this in tables 3 and 4.

[1] James had written asking for a copy of Fisher's 1941 paper, 'The asymptotic approach to Behrens' integral, with tables for the *d*-test of significance' (*CP* 181).

Fisher to G.S. James: 16 *August* 1955

Thank you for your long letter, which, however, broaches many topics beyond those of which I wrote. Returning to the content of my letter of August 13th, I gather you do believe that the strength of the significance under which a composite hypothesis can be rejected is to be identified with the frequency with which such evidence is obtainable.

I had thought that few other than Egon Pearson accepted this view, to which there are a great many counter-examples. It would therefore be useless to argue the matter. However, with respect to what you say on fiducial probability, I see that you have been misled partly by early misapprehensions of my own, because for a time I imagined that the logical content of a statement of fiducial probability could be distinguished from that of a classical probability, in the sense of Bayes and the early writers. I had also not sufficiently emphasized that fiducial probabilities can be demonstrated by an entirely rigorous step-by-step argument which excludes all possible ambiguities. However, I am averse to controversy in print or in letters.

Fisher to H. Jeffreys: 23 *February* 1934

I understand that the Royal Society are prepared to publish two articles from yourself and from me, on the Principles of Inductive Inference, on condition that we see each other's contributions before the articles are accepted. I enclose a short note[1] which gave rise to this decision, which embodies the arguments that weigh with me on some of the topics that may be raised. I should be glad to have it back for a revision with the first draft of your contribution.

[1] Eventually published as *CP* 109.

H. Jeffreys to Fisher: 24 *February* 1934

Thanks for loan of your paper. I have glanced through it but of course have not yet had time to make up my mind about all the points. However my main reaction is that the difference between you and me on any questions that concern you directly is not great. Wrinch and I showed in our original *Phil.*

Mag. paper[1] that in sampling problems, provided the sample is large, that for a wide range of distributions of the prior probability the posterior probability is practically proportional to the p. of the sample given the whole class, and therefore to your 'likelihood'. Consequently allowance for prior probability in these matters makes little difference except in freak cases, as when the sample consists of five white balls in a bag we knew to contain 100 in all of which just 5 were known to be white. Your expressions in ordinary cases are good approximations to the posterior probability; if I claim in one case to have got a slightly better one I don't see that you need worry. If in your original *Phil. Trans.* [*CP* 18] you had referred to this result of ours as providing an alternative justification of your results you would have been on strong ground. I notice that in your paper you don't defend your statement that the ratio of two infinite numbers has a definite value; and this is my crucial objection to your point of view.

My point is that we need a theory that will allow us to draw inferences in cases where sampling methods do not apply, and it must have two properties. It must make it possible to assimilate new experimental knowledge; and it must give us ground for believing the laws that we do believe or else say definitely that our inferences are fallacious. You would accept the first; but for the second you must introduce *a priori* considerations of some kind. A quantitative physical law is on an entirely different footing from any type of sampling, for reasons given in the Quantitative Laws chapter of my *Scientific Inference*.[2] It would be a quite logical position to maintain that all quantitative laws are merely interpolations and that all values of y for other than the observed values of x are equally likely; e.g. that the predictions in the *Nautical Almanac* are meaningless; but if you really mean that, I think you ought to say so.

My quarrel is not with you, but with Eddington and similar people. E. would say that the law of gravitation is determined by purely *a priori* considerations; I say that it is partly *a priori* and partly experimental. You would deny the *a priori* considerations altogether. So it is rather anomalous that in a recent Phys. Soc. paper on measurement E. throws as much mud as he can at probability (while introducing probability considerations at every turn without noticing them) and at the same time quotes some of your work with approval. Such an attitude means a complete *a priori* theory and an experimental one, with no possible point of contact. I prefer to have just enough *a priori* material to give experiment a chance to decide what is true. You give too little; Eddington gives too much.

I cannot follow your objection to the generalization to all probabilities of the laws of probability obtained for samples. You do not deal with my argument for the need for an *a priori* postulate, and your argument would say equally well that we cannot determine the distance of a star from its parallax.

[1] Wrinch, D. and Jeffreys, H. (1919). On some aspects of the theory of probability. *Phil. Mag.* **38**, 715–31.
[2] Cambridge University Press, 1931.

Fisher to H. Jeffreys: 26 February 1934

Thanks for your letter. I am glad you think the differences in our points of view do not go so deep as one might judge, as it would be a pity if we occupied the space offered by the Royal in imitating the Kilkenny cats.

I myself feel no difficulty about the ratio of two quantities, both of which increase without limit, tending to a finite value, and think personally that this limiting ratio may be properly spoken of as the ratio of two infinite values when their mode of tending to infinity has been properly defined. I gave a rigorous statement on this point in the *Proc. Camb. Phil. Soc.*, at the beginning of a paper entitled 'Theory of Statistical Estimation' [*CP* 42]. The proposition there stated is not difficult to prove, and I cannot see that it leaves any ambiguity as to the meaning of frequency ratios in infinite populations. The question has been discussed in other terms by von Mises, but his definitions applied, I believe, only when the populations are denumerable—an unimportant but necessary restriction, seeing that we often use probabilities proportional to lengths or areas.

I do not object to the generalisation to all *probabilities* of the laws appropriate to the games of chance, but I do think, and indeed claim to have shown, that there are also logical situations in which a rigorous statement of the nature of uncertainty in our uncertain inferences is expressible not in terms of probability, but in terms of likelihood, a quantity which does not obey these laws. The derivation of probability statements from statements involving likelihood, in the special cases where such derivation is possible, interests me greatly, and seems the right starting point for exploring the almost unknown field of the relations between probability and likelihood.

Here is an example of the kind of problem in this connection which puzzles me. A man makes genetical tests on a number of plants from a wild population. He finds three of them may be called a, two more b, two more c, and one each of d, e, and f. His data thus consists of the partition (32^21^3) of the number 10.

Let him know or be willing to assume that different types occur in the population with frequencies in the ratio $1:r:r^2:r^3:\ldots$ For any value of r he can calculate the probability of getting his observed partition, supposing his sample has been chosen really at random. He thus knows the likelihood of all values of r, but unless he has prior knowledge as to the distribution of r, I think you will agree with me that he does not know the probability of r exceeding any assigned value. For each value of r there is, I think, a calculable probability that the next plant to be tested will be of a type not hitherto found, and this probability will, I suppose, increase from 0 to 1 as r increases from 0 to 1. What sort of information has he then about this probability?

You would, I think, approach the problem introducing some sort of prior knowledge, which would make this probability definite though it would depend on the prior knowledge introduced, but you would not be unwilling to conceive that a rational being might happen to lack the prior knowledge of the kind introduced, and yet you would, I suppose, be unwilling to assert that no amount of experience without such prior knowledge could give him any guidance as to whether to expect new types or not. It looks as though some sorts of rational inference require both the concepts of probability and likelihood in a rigorous statement of the nature of their uncertainty.

H. Jeffreys to Fisher: 1 *March* 1934

Thanks for your letter. I think that even if we can't agree and have to leave a lot for other people to argue about, we ought to be able to agree about what our points of difference are. At present it seems that the relations between us are of the form 'A thinks B has done all the things that B has been at special trouble to avoid'; where if A = J, B = F and conversely. With me might I think be included F.P. Ramsey and C.D. Broad; there are two very important papers by the latter in *Mind*, vols. 27 and 29,[1] which I read when they came out, and they seem (especially the first) to have influenced me to the extent that I assimilated the ideas and forgot where they came from. I'll have to make a belated acknowledgement somewhere.

On the question of ratios of infinities first. In your *Phil. Trans.* [*CP* 18] you quite explicitly excluded the notion of a limit and stated the definition of a probability as the ratio of two infinite numbers. The limit was Venn's dodge, and you dropped it. But when it comes to the point neither you nor Venn use your own definitions. You have never taken an infinite class and counted the number of a sub-class within it, nor has Venn ever found a probability as the limit of a ratio when the number of trials tends to infinity. (By the way, you use the word 'probability' in your definition; but later on you generally use 'frequency'. I suppose you mean the same thing by both, but am not sure.) When you want actual numerical estimates you assess directly the frequency of a sample in the form ${}^rC_\ell\,{}^{n-r}C_{m-\ell}/{}^nC_r$. So far as I can see this is got by counting cases in a perfectly correct way and then saying that all cases are equally probable. On the *a priori* view of probability there is no more to be said; on the frequency view it may be right or not, but the point I want to make is that you don't avoid making an *a priori* assumption. You have to assume that if you take an indefinitely large number of classes of number n, each of which contains r ϕ's, and sample them, the ratios of compositions of the samples will occur in just the frequency given by your estimated probability. You do not know this by experience and therefore it is an *a priori* postulate. You may think that it is more plausible than the direct use of non-sufficient reason; but that is not a reason for condemning the latter on the ground that it is not known by experience. So far as I can see your practical results (in the *Phil. Trans.* I mean; I have not followed your later work in much detail) can be taken over into the *a priori* theory of probability without change of their quantitative statement, though some of them may need a change of language.

To take another case; suppose that by your methods you compare two methods of growing potatoes and show that in 90 per cent of cases method A will give a greater yield than method B. Suppose a farmer asks you 'What reason is there to suppose that *I* will get a greater yield by method A?' It seems to me that your only answer is in terms of the view that probability is intelligible without definition.

There is a criticism of the *a priori* view that I often meet and have never succeeded in understanding, and you give me some clue to it at last. It is the fundamental objection to the idea that $P(p|q)$ has a definite value whatever p and q may be. I am getting inclined to think that behind this objection is an idea that I think that q is *relevant* to the truth of p in all cases, which is another matter altogether. In the majority of cases we have things like $P(p|qr) = P(p|q)$ whatever r may be over a very wide range, i.e. in words, r is irrelevant to the truth of p given q. E.g. on my present knowledge of your movements it is as likely as not that you will be in Cambridge before you are at Rothamsted; the probability is $\frac{1}{2}$. If I find that you still have a house at Rothamsted the probability sinks nearly to 0; but it is only such quite special additions to information that will materially alter the probability.

You also seem to suggest in your paper that when I assess a prior probability I am expressing an opinion on the ratio of the numbers of cases in the world. My attitude, on the contrary, is that it is only on my view that it is possible to make any progress without expressing such an opinion. I started from the pure phenomenalist position, but found that phenomenalism needs a good deal of amplification before it can deal with the problem of inference. We don't know all about the world to start with; our knowledge by experience consists simply of a rather scattered lot of sensations, and we cannot get any further without some *a priori* postulates. My problem is to get these stated as clearly as possible. The tests available cannot be experimental; the conditions required, in my view, are that we want just enough *a priori* hypothesis to make it possible to settle the rest by experience. We cannot exclude logically the possibility that when a law looks well established the next attempt at verification may jigger it up altogether, so that general laws are never certain at any stage but only have a certain degree of probability. I got as far as I could with the principle of non-sufficient reason, but it turns out in some cases to give answers quite contrary to general belief. E.g. as Broad pointed out, it will never give a reasonable probability to a general law of the form 'all crows are black'. He tried to get a workable alternative but got no answer very convincing either to him or to me; I discussed the point very shortly in *Scientific Inference* 191–197, but I think the attempt in my *Camb. Phil. Soc.* paper on sampling[2] is on the right lines. The case of quantitative laws is even more to the point. E.g. in the case of uniform acceleration discussed in *Scientific Inference* 37–41, we could choose an infinite number of laws that would fit the data perfectly, but nevertheless do choose one, on *a priori* grounds, that only fits them approximately. If likelihood was the only thing that mattered we would choose any of the exact solutions in preference to the one we actually do choose.

The principle of non-sufficient reason is intended to serve simply as an expression of lack of prejudice; in a sampling problem we want to give all constitutions of the whole class an equal chance of acquiring a high probability by experiment. But in these cases of general laws there seems to be prejudice; I cannot help it, but there is a general belief in the possibility of establishing quantitative laws by experience, and I am not prepared to say that the general belief is wrong. I think I have stated a postulate that expresses it sufficiently clearly for practical purposes, but the postulate cannot be proved experimentally. But since such a postulate must from its nature be believed independently of experience that is a recommendation. What I want is, since an *a priori* postulate is needed anyhow, to choose it in such a way that the maximum number of alternatives are left for experience to select from. Eddington, e.g., chooses it too drastically by saying that the law of gravitation *must* be generally covariant and that space *must* have a finite curvature, deliberately excluding 0 from the admissible values of a certain coefficient.

Your paper on Inverse Probability in the *C.P.S. Proc.* [*CP* 95] involves a subtle point. Suppose we have a huge population and from it make up classes each of 10 000 members, containing respectively 0,1,2 . . . with the property ϕ. We choose one of these at random and select a sample of 100, of which 30 have the property ϕ. You say, and I agree, that this establishes a high probability that the class chosen contains about 3 000 with the property. If on the other hand we choose 10000 at random, sampled them as before with the same result, we should still estimate by inverse probability, using n.s.r., that the class contained about 3000 ϕ's. You think it absurd that the two should give the same result. Actually they do when you only sample one class; but if you take two classes there is a difference. In the first case the fact that one class has been estimated to contain about 3000 ϕ's slightly reduces the probability that the second will contain a number in this range, because one of them is excluded from the possible compositions of the second. But in the second case the first sample of 100 is effectively a sample of the whole super-population, and establishes that the ratio in this is probably about 3 to 7. Hence there is an *increased* prior probability that the second 10000 sampled will be in about this ratio. If we again get 30, we shall feel increased confidence that the number is about 3000 (and reduce the probable error); but in the first case what change there is, is in the other direction.

I don't understand your remark about a verbal point in your criticism of my remarks on Keynes. Mises said something similar in reviewing my book. My immediate reaction is that I cannot take it as obvious that people's linguistic habits are meaningless; but I think there is something else and can't see what it is.

I keep on getting other jobs pushed on to me and have not yet managed to read your paper carefully, though I have re-read a good deal of previous work. . . . In your discussion of my 1/3 business [*CP* 102] you seemed to assume that the third observation was equally distributed in some way; this is contrary to the postulate that the probability follows the normal law. I don't think there is much to add to Bartlett's discussion. My own got rather muddled through getting statements in the wrong order. The point is that if we know anything about h to begin with, and the first two observations are separated by a big multiple of $1/h$, there will be an extra probability that the third will lie between them; if they are separated by a small multiple, it will be probable that the third will lie outside them by a distance of order $1/h$. When the first two are fixed, the probability that the third will be between them is 1/3 only if we have no previous knowledge about h.

I will think about your problem of the trees.

I think that as a result of this you may feel inclined to alter your paper somewhat; if I answer it as it stands I think I shall miss the important points and say too much about others. Should I send it back for any modifications you think desirable and then I can have another try?

[1] Broad, C.D. (1918). On the relation between induction and probability—(part I). *Mind* **27**, 389–404. Broad, C.D. (1920). The relation between induction and probability—(part II). *Mind* **29**, 11–45.

[2] Jeffreys, H. (1932). On the prior probability in the theory of sampling. *Proc. Camb. Phil. Soc.* **29**, 83–87.

H. Jeffreys to Fisher: 21 *March* 1934

I have got this more or less done at last; sorry to be so long. I have not typed it as there may be further revision.

I think it may help readers if we could put things the other way round. I have arranged my paper so that it will be an explanation of my own point of view; how far the opinions I have dealt with are yours I don't know, but as they are some people's I think they should be discussed. I think your paper is not relevant to my real arguments and that as it stands it will tend to obscure things; so I suggest that mine might come first and you might add comments at the end. That of course is as you like. I could put a note at the beginning to day that this is the result of discussion between us.

I have thought a bit more about your question about the wild plants, but am hazy about what you want. If it is a sampling problem, total number of individuals n, with r_1, r_2 . . . of the various species, I think that $P(r_1, r_2, \ldots|n)$ should be constant for all values subject to the sum being n. This seems the obvious generalization of Laplace. W.E. Johnson had a paper in *Mind*[1] about a year ago (posthumous, edited by Braithwaite) in which he did something of the sort; the remarkable thing about it was that he did the whole thing on the posterior probability, but he had an undetermined constant left over. That's the trouble about philosophers; they talk sense up to a point (some of them I mean) but run away when they're on the verge of saying something useful. On the other hand you may be thinking of a Mendelian problem where there is one species and the question is about the incidence of one factor. Then I think my *C.P.S.* paper is relevant. But I may have missed the point altogether because I couldn't see where your ratios $1,r,r^2$ . . . could have come from.

Frank Ramsey seems to have had the gift of thinking clearly about these things more than anybody I know, though I have hopes of Broad. If somebody of that type would get on to it I think we could make some progress. I'm not desperately interested myself, being willing to take approximations and even orders of magnitude when it is too much trouble to get anything better; but if e.g. the simplicity postulate could be stated more precisely than I have done it would serve as a better guide than we have yet as to when we ought to introduce a complication. The sort of thing that bothers me is this. In seismology we get times of transmission to various distances, and fit a polynomial of degree 3, say, to them. The significance of the last term really involves the prior probability that such a term will be present. The usual thing is to keep it if it is some arbitrary multiple of its standard error, but I think it ought to be possible to frame a rule with *some* sort of argument behind it. . . .

[1] Johnson, W.E. (1932). Probability: the deductive and inductive problems. *Mind* 31, 409–23.

Fisher to H. Jeffreys: 29 *March* 1934

Thanks for your letter. We seem to have a deal of ground to cover in our ten pages apiece. I think your draft shows very well what your line of approach is; at least, I find it much more reasonable in this form than I did in the *Proceedings*.[1] I want rather to modify the tone of my note, without much altering its substance, e.g. instead of saying that in your definition notions analogous to probability and those analogous to amount of information are 'confused', I have altered the word to 'admissible' [*CP* 109, p. 6.].

From the point of view of interesting the general scientific public, which really ought to be much more interested than it is in the problem of inductive inference, probably the most useful thing we could do would be to take one or more specific puzzles and show what our respective methods made of them. This might mean more work than you want to undertake and will certainly require that we should agree quite exactly as to what the data were and what questions we want to elucidate from them. I had this in mind in mentioning the plant problem, as it was the first case I had come across in which the notions of probability and likelihood appeared to be harnessed tandem, instead of only one or the other of them.

In case you feel it is a point worth discussing I will state it again in a form

which, apart from complications, which I think are logically unessential, is that in which it might arise in practice.

A man collects a number of plants of the same species from a wild population, grows them in experimental cultures, and after some years of much labour, has succeeded in classifying a small number. Three of those tested he finds to be identical in respect to his method of classification and these he calls type *A*. Two more differ from *A*, but agree with each other. He calls these *B*. Two more differ from *A* and *B*, but being alike, constitute a new type *C*. The remaining three which he has tested differ from *A*, *B* and *C*, and from each other, and so constitute solitary representatives of types *D*, *E* and *F*. He has then tested ten plants and found them to belong to six different types with frequencies of occurrence represented by the partition (32^21^3) of the number 10.

He knows, or thinks he knows, that the frequencies with which the different sorts of plants distinguishable by his tests occur in the wild population constitute a geometric progression like the frequencies of Planck's oscillators containing 0,1,2,3,. . . quanta of energy. He is not concerned to test this hypothesis, but using it as part of his data to estimate the probable outcome, in respect of the discovery of new types, of continuing his labours by testing more specimens from the wild population, the kind of question he might properly ask is, 'What is the probability that the eleventh plant tested will turn out to belong to one of the six groups already found, or alternatively to add a new type to the list?'

If he knew the value of r, the constant ratio of his geometric progression, this probability would, I think, be determinate and for different values of r from 0 to 1, the probability of getting a new type next time must also, I suppose, run from 0 to 1. He does not, in fact, know r, but his observed partition gives him some information about r, e.g. if the observed partition had been (1^{10}) he would think r was higher than he would if the observed partition had been (10). He can in fact assign to each value of r a certain likelihood, which will be higher if his partition is made up of a lot of little parts and lower if it has a few big parts, and for any observed partition he can specify the particular value of r which is most likely and the likelihood relative to it of any other value of r.

But this takes me no further than to tell the botanist that I know which of the possible values of the probability he is seeking has the highest likelihood, and I know how much proportionately the likelihood is lower for any other value of the probability. The real question is whether he has any right to expect from me more definite information than this and this may be considered under the heads (a) are any further deductions to be drawn rigorously from the bad data as stated? (b) can we *properly* supplement these data by axiomatic truths relevant to our problem, which will enable further conclusions to be drawn?

So much for the case that puzzles me. Without suggesting that you should modify your draft, it may be worthwhile for me to make a few comments. You say I object to the introduction of an *a priori* element, [so] I should like to get this notion a little more precise. What I object to is the assertion that we must, in considering the possible values of an unknown parameter used to specify the population sampled, introduce a frequency distribution or a probability distribution of this parameter, supposingly known *a priori*. If we really have knowledge of this kind, as in some problems, which can be reconstructed with dice or urns, I do not deny that it should be introduced, but I say that we often do not possess this knowledge, i.e. either that we are in absolute ignorance or that our knowledge is so vague or unsatisfactory that we may properly prefer not to introduce it into the basis of a mathematical discussion, but rather prefer to keep it in reserve and see what the observations can prove without it and whether it, for what it is worth, is confirmed or contradicted by the observations. I claim, in fact, that it is at least a legitimate question to ask: 'What can the observations tell us when we know the form of our population, but know nothing of the *a priori* probability of the different values that its parameters may have?' This is the situation which I treat as the typical one in the Theory of Estimation, but it would be quite legitimate to say that our assumed knowledge of the form of the distribution, which is needed before there can be anything to estimate, is in the nature of *a priori* knowledge. In fact, such knowledge seems to me essential in what I call problems of specification, but out of place in the next stage, when problems of estimation arise.

I think this may explain your difficulty about fitting a smooth curve to a number of discrepant observations. If the form of the curve is given, e.g. a polynomial of the fourth degree, then you will probably agree that the best curve of this form will be one specified by maximum likelihood. This would not prevent one from concluding, if after much experience it were found that some simplification of the formula were possible, e.g. that the coefficient of the fourth power seldom differed significantly from the difference between the coefficients of the second and third powers, that our specification could be revised with advantage, but while any specification is in use the problem of estimation can be made perfectly definite and it is primarily from the advantage of separating all the complications of the theory of estimation from the logical difficulties of specification, that I prefer to think of the thing in these two stages. In practice, as we know, simplicity is sought in specification, but whether it is to be justified primarily by convenience, subject to the satisfaction of tests of goodness of fit or whether, as you have proposed, it should more properly be justified in terms of probability, I have not any strong opinion. If our object were to explain the habits of a group of organisms, known as scientific men, I do not see that convenience is a less effective explanation, though certainly a less distinguished one than probability.

This is enough and more for one letter. I am returning the drafts.

[1] See Fisher's letter of 26 September 1933 to Bartlett (p.46).

H. Jeffreys to Fisher: 10 *April* 1934

I find your genetics problem rather hard. I think it might be tackled like this. Suppose you have a population of one type and breed from them; and that the probability of any new one showing a mutation is r. Then I should say that the prior probability of r is uniformly distributed if you know nothing about r to begin with; i.e. if $f(r)$ dr is the p.p. [prior probability] that r is in [the] range dr, $f(r)=1$. More accurately perhaps, if you are testing the occurrence of mutation in a particular factor, the prior p. that $r=0$ is 1/3, $r=1$ is 1/3, and $f(r)=1/3$ otherwise. For another factor the same values; and as far as we know to begin with the two are independent so that $f(r_1, r_2)\mathrm{d}r_1\mathrm{d}r_2 = f(r_1)\mathrm{d}r_1 \times f(r_2)\mathrm{d}r_2$. I can't see what would make your biologist think that r is the same for all factors, but if he is sure of it then the same $f(r)$ holds for all. I think that if he knows this he must know a lot of other things too and this form may be quite wrong. We might have to allow for our knowledge that in fact most things breed true, which would concentrate $f(r)$ towards the lower values; but I think the above is right if he really knows nothing.

Newman once asked me the following: a man arrives at a railway junction in a town in a foreign country, which he has never heard of before. The first thing he sees is a tramcar numbered 100. Can he infer anything about the number of tramcars in the town? Newman thought the question was significant and so did I, and we both had a feeling that there were probably about 200. I tried it on M.S. Bartlett, who thought it was meaningless but had the same feeling about 200. This may have some analogy with the question of the existence of a further factor liable to mutation. My very doubtful solution is: let $f(n)$ be the prior probability that there are n cars; the probability that number 100 would be observed, given n, is $1/n$ for $n \geqslant 100$, 0 for $n<100$. Hence the posterior probability of n is $\propto f(n)/n$. If $f(n)$ is constant $\Sigma f(n)/n$ diverges and we can infer nothing (or rather, that it is practically certain that $n>200$!); but if $f(n) \propto 1/n$, $f(n)/n \propto 1/n^2$ and the probability that n exceeds m (>100) is

$$\sum_m^\infty \frac{1}{n^2} \Big/ \sum_{100}^\infty \frac{1}{n^2} \approx \frac{1}{m} \div \frac{1}{100}$$

and it is about as likely as not that $m>200$.

I have just read Bayes's paper again for the first time since 1919. I see he defines probability in terms of expectation of value, as Ramsey does; R. has really given only conditions for consistency. Bayes gets $P(pq|h) = P(p|h) \times P(q|ph)$ quite correctly; I can understand him rather better than Ramsey but that is probably my fault. What bothered me about R. was that I did not see that people would agree in their ordering of values any better than in their ordering of probabilities; it really involves an ideal man who would always act on the alternative that gives him the greatest expectation of value. But it's really no worse than mathematics, which really assumes an ideal man that always gets his arithmetic right.

Towards the end of your letter you speak again of unknown *a priori* probabilities as opposed to those arising with dice. My attitude is that you are asking the wrong question. My question is, what distribution of probability corresponds to absence of previous knowledge? I think you are regarding a probability as a statement about the composition of the world as a whole, which it is not and on a scientific procedure could not be until there was nothing more to do.

I have just been away cycling for ten days.

Fisher to H. Jeffreys: 13 *April* 1934

Thanks for your letter of April 10th. You have not got hold of my botanical problem yet, for it has nothing to do with mutations or mutation rates, but solely with the proportions in which different types, which can be recognised by experimental breeding as distinct, exist in a wild population, i.e. for any two plants the experimenter can, in time, ascertain whether they are alike or not alike, and so can find that he has 6 different sorts in the 10 plants tested. But he knows nothing about which is the mutant type and which the parent type.

The information which I supposed him to have, that the frequencies of different types in nature were in geometric progression, might have been different without altering the logical form of the problem, e.g. an alternative problem might have been to interpret the same data on the supposition that there was in nature a finite but unknown number of types, all equally frequent. One could then infer the likelihood of any proposed number as great as or greater than the number so far found and so the likelihood of each of the possible values of the probability of getting a new type at the next trial. Either item of supplementary information leads to essentially the same logical situation though with a different mathematical content.

I like Bayes' approach to the notion of probability, because it is historically true that expectations, as of cargoes at sea or contingent rights in property, must have been bought and sold long before the value placed upon them was formally analysed into the two components, probability of possession and value if possessed. Indeed it is curious, though I believe it is a fact, that this analysis was not carried out by the Greek or Arabian mathematicians or indeed by anyone until the sixteenth century, if then. It seems to me the basis for what we may call the frequency aspect of the probability theory.

The tram problem is a good one. If instead of a whole number, you have a continuous variate, e.g. if the traveller arrives by parachute near a stone marked 'City of X, 1 km. from city centre', he has quite a nice basis for estimating the radius of the city supposing, of course, that he knows that all the cities are circular and that distances are marked from their geometrical centres. He then has an unambiguous fiducial argument as follows. 'If the radius of the city exceeds R kms., the probability of falling on or within the 1 km. circle is [less than] $1/R^2$. If this event has happened the fiducial probability that the radius exceeds R kilometres is therefore $1/R^2$ and the fiducial probability of it lying in the range dR is $2dR/R^3$. The fiducial median city has a radius $\sqrt{2}$ kilometres and an area 2π. The fiducial mean radius is 2km. and the fiducial mean area is infinite. The most likely [radius] for the city however, is 1km., for this maximises the probability of his observation.

I don't think this quite applies to the tram-cars. If No. A has been observed one can argue that if n is the total number the probability of meeting with the number A or less is A/n and if n exceeds any value N_0 this probability is less than A/N_0. But one clearly cannot differentiate and at first sight I do not see how to handle the inequalities. In fact I fancy that the reason why the fiducial type of argument was for so long unrecognised is just that the earlier writers always took frequencies, that is discontinuous observations, rather than measurements as the typical data in the theory of probabilities. Continuous variates with their standard deviations and regression coefficients are, in this matter, susceptible to a much simpler rigorous treatment.

H. Jeffreys to Fisher: 20 *April* 1934

I've got this typed at last. I have rearranged and altered it a bit. Would you mind filling in the reference to a paper of yours on p.9? . . .

Fisher to H. Jeffreys: 26 *April* 1934

I am sending you back the type-scripts, so that you can see my last alterations, at least I hope they may be the last. If you think they will do, will you send them both in to the Royal, as they are agreed documents. I think, on the whole, the suggestion of the Mathematical Committee[1] has been justified and that our notes, within the limits of space assigned, will be of more general interest than if we had written independently.

[1] See Fisher's letter of 23 February 1934 (p.149).

H. Jeffreys to Fisher: 9 *May* 1937

A question has just arisen about the excess motion of the node of Venus. It is 3.5 times the standard error, the probability of a random deviation exceeding which is 0.00041. Eddington says that as it is one of 15 it can be accepted as normal. The p. that one of 15 would exceed 3.5σ is 0.006. What I should like to know from you is whether there is another case on record where a statistician has accepted at sight a deviation beyond your 1% limit as random? (The other 14 give a χ^2 of 15).

By my test the thing is probably random on account of the large number of observations combined, but there's not much to spare, and the situation would be altered if some *specific* systematic error was before the House.

I should be grateful if you would send me some of your papers. I'm not much interested in agriculture but I should particularly like (1) the one where you got the distribution of X^2, (2) the one where you introduced z, (3) one that Wishart mentioned to me the other day about X being distributed nearly normally.

Fisher to H.Jeffreys: 11 *May* 1937

I am sending a few mathematical offprints that may, I hope, be of interest to you. The problem you set me is clearly one of judgement, and not of calculation, and indeed you seem to have done all the calculations needed. I should be inclined, naturally, to accept Eddington's judgement on an astronomical point, especially as your own test seems to confirm it. On the other hand, *prima facie*, i.e. on an assumption ordinarily made, the probability 0.006 is amply small enough to claim significance, and would be used for this purpose with complete confidence, I have no doubt, if anyone had a theory which required such a deviation.

H. Jeffreys to Fisher: 18 *May* 1937

Thank you very much for your letter and papers. Your letter confirms my previous impression that it would only be once in a blue moon that we would disagree about the inference to be drawn in any particular case, and that in the exceptional cases we would both be a bit doubtful. I am particularly glad to note your treatment of an unknown systematic error as a motion not before the House, while a specified one can be treated on its merits. This is what I have been saying about the 'excluded middle' for a long time.

Your letter encourages me to ask whether you also agree with me about the ideas of Eddington and Milne that all valid observations can be deduced by pure thought, without reference to other observations. As far as I am concerned this opinion does not agree with experience, and I prefer the alternative explanations that E. and M. are either very skilful in selecting their observations or very lucky. I agree completely on this with Dingle's recent *Nature* article.[1]

It was my disagreement with Eddington over this matter that first brought me into the subject in 1919, and I have been trying at intervals ever since to do for induction what Whitehead and Russell did for deduction in the introduction to *Principia Mathematica*: to get the postulates down to a minimum. You are doing the same. My only objection to your treatment of estimation is in the infinite population, which you introduce only in order to give a meaning to probability, while recognizing that it is hypothetical. I recognize that probability is hypothetical, so that the number of hypotheses is the same. It only happens that I find my way easier to understand. Any theory of induction is bound to be a bit fuzzy at the edges, but apparently you recognize

that too, though you put the fuzziness at the fiducial limits, whereas I put it into the practical difficulty of assessing the prior probability when the previous data are abundant but inadequately classified.

I am writing this because there is a tendency about to attribute what I believe to be an entirely exaggerated idea of our disagreement to us, for which we are both possibly partly responsible, and I think an occasional mention of cases where we agree would be for the good of the subject. I have a paper coming out soon in the *Proc. Roy. Soc.*,[2] partly on what had seemed to me the entirely mysterious fact that 'Student's' z distribution and mine obtained by the inverse theory are identical; there turns out to be a perfectly genuine reason why they should be the same in form, and with a trifling extra assumption 'Student's result can be made into an alternative proof of mine.

[1] Dingle, H. (1937). Modern Aristotelianism. *Nature* **139**, 784–6.
[2] Jeffreys, H. (1937). On the relation between direct and inverse methods in statistics. *Proc. R. Soc.* A **160**, 325–48.

Fisher to H. Jeffreys: 19 *May* 1937

I have been a great deal interested in Milne's recent writings, and have been a little puzzled by, I think, the same point as you, namely, his taste for developing the subject as much *a priori* as possible. I do not think he always felt like this about it; indeed his first reaction to Eddington's *a priori* approach was, if I remember right, rather that the interesting thing about the universe was that we had to look at it and see how it worked, instead of inventing it by introspection. It may be now that he has been working on some general and reasonable postulate which he has not yet made fully explicit, such as, that the observer has a right to assume that he is not too exceptionally placed in respect of the observations he can make, or that the laws of motion must be expressible in terms of relative only. I do think that his dynamical work is a marvellous achievement, and it seems at present to be in the way of clearing up the whole tangle for which Einstein and his following are responsible; but, much as I am impressed by his work, I am afraid I have not followed it closely enough to give an opinion of real value. . . .

H. Jeffreys to Fisher: 20 *May* 1937

Milne's general principle is something to be found in a part of Mach's work that I possibly never read, and from his earlier work I should be surprised if Mach thought it important: that as mass can be inferred only from relative motion, the mass of any body is caused by the motions of other bodies. It looks to me like a lapse into a metaphysical idea of causality that Mach had already disposed of. But unfortunately in order to get a frame of reference Milne has to assume a substratum with a finite density, tending to infinity at an outer boundary, which is itself spreading out with the velocity of light; and then he assumes that motions with respect to this substratum are small. As the substratum is unobservable I don't see that he has any justification for the second assumption. I should begin to be interested if he explained the things where the velocity of light and gravitation appear to be mixed—the three Einstein crucial tests. Judging simply by capacity for predicting observations I should prefer Eddington to Milne. On philosophical outlook my order of preference would be Einstein first, Eddington and Milne nowhere. Einstein has, I think, always made it clear that he is making suggestions that look hopeful and working out consequences that may lead to a test.

I am sending you a few early papers of mine that you won't find mentioned in any of the official relativity books; but they constitute my reason for taking the whole business seriously. . . .

H. Jeffreys to Fisher: 5 June 1937

. . . I have asked the Press to send you a copy of the new *Sci. Inf.* with the addenda. I should really have liked to scrap the lot and do it again, but at the present rate it looked as if the thing would take about 60 years to sell out. Looking at it again I don't seem to have made enough of the places where I agree with you, but I think I have remedied that in an article I have sent to *Nature* for the Dingle controversy.[1]

When I have time I should like to write a note showing how some of your methods fit into my theory. Maximum likelihood is obvious (I can't understand why K.P. was sticky about this, since m.l. for a large number of observations is an immediate consequence of inverse probability, and K.P. always accepted inverse probability) but sufficient and efficient statistics also have a definite place. The situation in my method indeed is that it goes straight for the efficient statistics and the mathematical difficulty comes in when we have to decide what to do when somebody has provided only inefficient ones.

[1] Jeffreys, H. (1937). Physical science and philosophy. *Nature* 139, 1004–5.

H. Jeffreys to Fisher: 30 *July* 1937

I have read your paper [*CP* 149] on the methods (plural intentional) of moments with interest. A bit on the severe side, perhaps, but probably called for. However you tempt me to ask whether you would consider a short paper for your journal on the relation of maximum likelihood to inverse probability. An idea seems to have grown up that they are opposed, and I think this is very much in need of correction. It is most odd that K.P. always accepted i.p. and yet was sticky about likelihood, and adopted a method of fitting that was inconsistent with i.p. . Wrinch and I gave the essence of the relation, for sampling, in our 1919 paper,[1] but it would stand a bit more generalization. We might have gone further at the time, but for one thing we thought it was obvious, and for another we thought that statisticians already used it; as

indeed they did for sampling and fitting normal distributions.

I am interested in your doubts about Pearson curves as such. They have not come my way much; I once got a distribution of the $x^p\, e^{-ax}dx$ form, in radioactivities of rocks, but have not had anything more complicated. In seismology the law of error is practically a normal law with a uniform distribution superposed; the infinite Pearson curves don't seem to fit the outlying observations very well, but it seems possible to locate the normal distribution very well without anything more complicated than the dodge I introduced in my Strasbourg paper, which is quite complicated enough when there are some hundreds of groups to fit. However I should like to know whether there is any evidence (1) that the K.P. function with infinite tails both ways is better than the exponential of a quartic, which would make the method of moments right, (2) whether the one with finite limits $-a$ and $+b$ is any better than is got by taking a new variable $y = \log\{(x+a)/(b-x)\}$ and fitting a normal distribution? The functions would no longer look like a single family, but I don't think that is any great objection. The K.P. ones aren't really because an imaginary tail is rejected to keep the results real.

In astronomy the question of correcting an observed frequency distribution for a known standard error is continually coming up. Eddington gave a method a long time ago (reproduced in Brunt) which is mathematically equivalent to solving the heat conduction equation for negative time. If he had the true probabilities presumably it would be all right, but it is applied to the observed numbers and if applied completely I think would necessarily diverge. Do you know any way of doing it otherwise than for the normal law? The kind of thing it is applied to is the distribution of absolute magnitudes of stars.

[1] See Jeffreys' letter of 24 February 1934 (p. 149).

Fisher to H. Jeffreys: 16 *August* 1937

I had to be rather down on Pearson, as for a time it seemed likely that his attack would wreck Koshal's statistical career, and I particularly wanted to make clear to our Indian brethren that unfair attacks on foreigners were as much resented in this country as they must be by their victims.

Like you, I have fairly often run into $x^p e^{-ax}dx$, but with the terminus at zero, and not at an arbitrary point as in the Pearson form. This makes the curve altogether more manageable.

I have long thought it a pity that the form with a quartic exponent had not been more fully studied. One or two attempts have been made, as by O'Toole in America, but it involves a transcendental function as interesting, I should think, as the hypergeometric which seems never to have been sorted out. I have very often found translation, i.e. functional transformation of the variate, extremely useful to make a distribution sufficiently or indeed nearly usually normal. I think, myself, it would be unreasonable to expect cases from all sources to conform to a single family.

The question you raise of removing a known amount of variance from a distribution is really a most intricate and intriguing one. I suppose the logical approach is by correcting the cumulants, but that, of course, only gives a finite series of the first few cumulants of the corrected curve, and to infer the true distribution from these not only sometimes involves approximations that fail to converge ultimately, but too often have a bad practical convergence of the terms available.

[P.S.] Do not hesitate to send along anything you would like to publish, but do not be offended if I think it unsuitable for the journal.

H. Jeffreys to Fisher: 25 August 1937

I enclose the paper[1] in question. My own feeling about it is that it rather labours the obvious, but as K.P. never saw the point, and as it came as a surprise to Yule when I mentioned it to him the other day, I think it will be unfamiliar to 90 per cent of statisticians. So I should be glad if you can find room for it.

I am not sure if I have got the right reference on p.4 about order less than $n^{-\frac{1}{2}}$; I have an impression that I have seen something by you that refers to the point more definitely, but a search has not revealed it.

[1] Eventually published as Jeffreys, H. (1938). Maximum likelihood, inverse probability and the method of moments. *Ann. Eugen.* 8, 146–51.

Fisher to H. Jeffreys: 8 September 1937

As I do not like the practice of editorial comments being made without the author's knowledge I am sending you the enclosed which I propose as an editorial note[1] on your paper. I hope and think that you will find it unobjectionable.

[1] See *Ann. Eugen.* 8, 151 (1938).

H. Jeffreys to Fisher: 5 March 1938

I heard yesterday that you were back; I hope you found India enjoyable. Meanwhile I have got out a number of things that may interest you. . . .

I read your *Design of Experiments* recently with much interest. It leads me to query your disparagement of your intuition in the note at the end of my paper in the *Annals*. At any rate wherever my methods, which are really based on the *Grammar of Science*,[1] lead to criteria differing from K.P.'s, you seem to have recommended something either identical with mine or so similar that hardly anybody would notice the difference. . . .

By the way where did you get the identification of a prior probability with a frequency in the external world? K.P.'s habit of calling his tables of probabilities frequencies suggests that he identified them; but I haven't found anything analogous in the work of anybody else that has used inverse probability. It is however always hard to be sure that a particular statement is not in Laplace somewhere.

[1] Pearson, K. (1892). *The grammar of science.* Black, London.

Fisher to H. Jeffreys: 7 March 1938

Many thanks for your interesting letter. I am very glad you like *The Design of Experiments*, as I had been concerned for some years previously with the kinds of logical difficulties that experimenters run into, and found it interesting and enjoyable to try and make an orderly statement of what can be done about them, putting the various dodges which have proved successful into their proper relations with each other. . . .

On your last point about identification of a prior probability with a frequency in the external world, I think I was quite a little influenced by Bayes, who, you may remember, defined probability as 'the ratio of the expectation to the value expected', and clearly regards expectation as a value averaged over objective frequencies of occurrence. I have always thought that this point of view must have been fairly widely spread in the 18th century, as it is the natural one for gamesters to take for such problems as evaluating expectations on unfinished games, but I do not know anyone who has brought it out so explicitly as Bayes.

H. Jeffreys to Fisher: 14 March 1938

I have just looked through Bayes's paper again, and I cannot see anything that suggests a frequency interpretation to me. He unfortunately does not give any examples of what he means by the value of an expectation, but I see nothing in his remarks inconsistent with Frank Ramsey's idea of comparing the desirability of different courses of action. (Ramsey does not appear to mention Bayes — or me — and he is as naive about causality as Russell; but he makes some good points). There would be little point in defining probability in terms of expectation if the latter was then going to be defined in terms of frequency — it would be simpler to define probability in terms of frequency straight away, and expectation would become a superfluous notion. Note too the remark on p. 410: 'It should be carefully remembered that these deductions suppose a previous total ignorance of nature'. This is, I think, clear enough that Bayes was not interpreting a prior probability as a statement of frequency.

The chief difficulty of Bayes's presentation is, I think, in his Prop. 1, where he assumes expectations additive. This needs a good deal more analysis, which Ramsey has supplied. I mean that the value of two dinners on successive nights may be additive, but that of two on the same night may be less than that of one, at any rate if the morning after is taken into account. So further restrictions are needed before the definition will work — it must be clear that the expectations added do not interfere with one another. I am not sure when the Weber–Fechner law was invented; if Bayes had known of it I think that he would have taken precautions; but personally I think it much simpler to take probability as a primitive idea and define expectation in terms of it when occasion arises.

The first definite identification of probability and frequency that I know is by R. Leslie Ellis, in *Camb. Phil. Trans.* 1843; Keynes [1] refers also to Cournot in the same year. Laplace in using the m/n ratio always sticks in the proviso that all the events must be equally probable, thus making probability a primitive idea, and this is also in Todhunter's *Algebra*, from which I got my first ideas. In a lot of later works this has been left out — e.g. by Neyman and Levy. Thus the Weldon–Pearson experiments on the bias of dice and the usefulness of Tippett's numbers as against personal choice to get a random sample become pointless — matters on which we should agree perfectly.

I notice that the same volume of the *Phil. Trans.* as contains Bayes's paper also contains a proof by him of the divergence of Stirling's series; I suppose this was the first.

I have been reading or rereading a good deal of mathematical logic during the last few weeks and am beginning to wonder whether a good deal of pure mathematics is not inductive. Whitehead and Russell, Ramsey and Wittgenstein all seem to agree that a number is a class of classes of distinguishable individuals, and therefore, one would think, that the existence of classes up to a certain number is an empirical proposition depending on the number of individuals in the world, which may well be finite. In that case there would be no such real number as $\sqrt{2}$, since the notion of a limit breaks down. The only possibilities seem to be that mathematics is either wholly idealist or that it is developed as science indicates that it may be wanted. Apparently even consistency cannot be proved. I sounded Littlewood about this and he shares my doubts. Hitherto I have regarded pure mathematics as fundamental and probability as using its rules, but essentially as a separate idea; but the alternative that pure mathematics is merely a special case of probability is going to need serious examination. At any rate on Wittgenstein's arguments it seems perfectly possible that the whole logic of infinity and therefore of limits and continuity is just bunk, and nobody can apparently see anything wrong with them. I don't see much prospect of solving the puzzle myself but I think that I may manage to limit the questions to a few definite alternatives.

By the way you said somewhere that the variation of observations from a formula is just as essential a part of the law as the formula itself, but I cannot trace the reference when I want it. It is a most important point from the logical and philosophical point of view.

[1] Keynes, J.M. (1921) *A treatise on probability*. Macmillan, London.

Fisher to H. Jeffreys: 24 *March* 1938

I have tried circularising your letter in my Laboratory with a view to discovering your reference to where I said that the variation of observations from a formula is just as essential a part of the law as the formula itself. Like you, I am quite sure that I have said so and constantly impressed it in lectures, but cannot lay my finger on it in print.

H. Jeffreys to Fisher: 17 *September* 1938

. . . I am wondering whether more attention shouldn't be given to the median. Its uncertainty seems much less sensitive to the form of the law than that of any other simple statistic, and significance tests for it are easy. The kind of case where it may arise is where there are only 8 or 10 observations and nothing much to indicate the character of the law. The mean of all or the mean of the extremes may have little value, but there's always something to be said for the median. I have some cases where the law is pretty certainly not normal, but whether it is $e^{-x/a}(x>0)$ or rectangular or $\{1-(x/a)\}$ $(0<x<a)$ there is no evidence whatever.

Fisher to H. Jeffreys: 19 *September* 1938

Thanks for your letter. I agree very much with your point about the median. It has the great advantage of having always an intelligible meaning, whatever the form of the curve. In writing Gosset's obituary for the forthcoming volume of the *Annals* I was interested to find that one of the first applications he made of his distribution was to apply it to a sample of two, and thereby get out an inference respecting the median irrespective of the form of curve. In writing the obituary [*CP* 165] I took the opportunity of developing this idea somewhat further, i.e. to any percentile point and to the simultaneous distribution of different percentile points. [1]. . .

[1] For Jeffreys' reply with comments on 'Student' and randomization, see his letter of 22 September 1938 (p. 270).

H. Jeffreys to Fisher: 20 *September* 1938

. . . The median of the *law* has an important position in my theory, because with any law, symmetrical or not, involving parameters of location and scale that are unknown initially, I find that the posterior probability, given the first two observations, that the median of the law lies between them, is always $\frac{1}{2}$. This is not true for any other parameter of location. For any other it is $<\frac{1}{2}$. (*Camb. Phil. Soc.* paper[1] of 1936, 423–5). 'Student' presumably got the analogous thing by his way of looking at it, which could be turned into mine by the argument I used in 'The relation between direct and inverse methods'.[2]

. . .

[1] Jeffreys, H. (1936). Further significance tests. *Proc. Camb. Phil. Soc.* **32**, 416–45.
[2] See Jeffreys' letter of 18 May 1937 (p. 162).

H. Jeffreys to Fisher: 30 *September* 1938

. . . If you are bringing out a new edition of *Statistical Methods* I wonder if you would modify the passages about inverse probability in it. I didn't agree with them when the book first appeared, but at that time i.p. fell between two stools. The logicians like W.E. Johnson, Frank Ramsey and Keynes accepted it, but did not develop it to any point where it was much use for actual application; a lot of others that accepted it also seem to have thought that the uniform prior p. of Bayes and Laplace was an essential part of the principle, and if this was so it would say that any estimated difference must be accepted — the null hypothesis would always be rejected. The practical men, in the circumstances, had a good deal of sense in not trying to use it in the state it was in. But it is now in such a state that it can be used. Either method would get the right answer some time or other; I think i.p. would give it, on the whole, more quickly, but as I still haven't found an actual case where our decisions would differ I haven't any strong views. The chief difference is about what order we should say the same things in.

I have been wishing for some time that a public benefactor would subsidize the C.U.P. to scrap *Scientific Inference* and give me a chance of writing something up to date. I should have written it in 1923, when Einstein was still news . . . .

Fisher to H. Jeffreys: 11 *October* 1938

I find I have left your letter of September 30 unanswered. My position about inverse probability is as follows: the term has been used, not always for quite the same thing, during the last 150 years, but fairly consistently to designate Laplace's extension of Bayes' method based on the concept of 'equal possibility', which I submit was used by Laplace merely to introduce a postulate of equal probability by a verbal subterfuge. A series of writers, notably in England, Boole, Venn and Chrystal, protested at different times against this hidden postulate, and even orthodox exponents, such as de Morgan, were evidently shaken in their faith by the arbitrary way in which the postulate was sometimes applied.

During the 17th and 18th centuries I think that mathematical probability was used consistently as a means of expressing expected frequency, but in the course of the controversy aroused by Laplace's doctrine a psychological view, well expressed in Keynes' phrase 'degree of rational belief', came to be advocated.

The opinion I express in *Statistical Methods* is, I think, quite clearly intended to refer to this old controversy, and not to the recent Cambridge work you refer to. I could, of course, in later editions — though I am afraid not in the 7th — add a sentence to say that I was not in particular discussing your theory of scientific inference, or your later work; but I really do not think that anyone yet has taken it to be so, but has only drawn the obvious

inference that you were attempting a resolution of the old enigma of a different kind from mine.

H. Jeffreys to Fisher: 12 *October* 1938

It is liable to be hard to follow old birds like Bayes and Laplace. People of that period often spent a chapter on a piece of mathematics that we should do in a page, and then slid over fundamental points with hardly a mention. Laplace's casualness about references in this case makes it harder. I think the real question is, did the critics of B. and L. from 1840 to 1900 or so understand them, or were they just attacking misinterpretations of them? In any case the misinterpretations have a long history, but they arose when B. and L. were not in a position to defend themselves. So far as I know K.P. was the only person that ever tried to *use* inverse probability and a frequency definition at the same time. I got my own first ideas from Todhunter's *Algebra*, the first relevant passage in it being practically a translation of Laplace, and it never occurred to me that any other interpretation than the one I make now could possibly be intended. I think it is relevant that I should regard a frequency definition of the prior probability as even more nonsensical than you do; you at least think it worth criticizing, whereas I should not think it worth mentioning if people didn't keep inflicting it on me. But if there was such a great difference between Bayes and Laplace and me as you suggest I should expect that reading them would make me feel sick, and it doesn't. I can't see anything in them that is not perfectly intelligible in terms of my own ideas, and I think the reasonable explanation is that they had the same fundamentals.

I don't agree that Laplace's 'provided all the cases are equally possible' is a verbal subterfuge; I should call it an indication that he had thought a bit more carefully than his predecessors. You strengthened my belief in this recently by referring me to de Moivre about something else. He starts off with a simple definition in terms of numbers of possible cases, as innocently as Neyman does. Laplace must have known of this, and the extra clause must mean that he saw that de M.'s definition didn't do what is needed. On the latter, so long as no face of a die is absolutely incapable of being thrown, the probability of any face is exactly 1/6 whatever other information we have. Laplace's allows for the possibility that if we get 50 sixes in 100 throws the p. of a 6 may be more than 1/6, even if all the others have actually occurred. Again, suppose we have two boxes, one containing a white ball and a black one, the other a white one and two black ones. Choose a box at random and then pick a ball at random out of it. What is the probability that the ball will be white? Answer, $\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{3} = \frac{5}{12}$. But on de Moivre's definition the assessment would be 2/5 — which I don't for a moment suppose he would actually say, but the only way of avoiding it is to recognize that the two white balls are not 'equally possible'. That is, even within the range of direct methods, a probability is not a mere matter of counting possible alternatives, and Laplace knew it.

Bayes's expectation theory is a bit muddled, because his rules correspond to what Laplace called mathematical expectation, but he is speaking of expectation of benefit, which is L's moral expectation. The distinction when the benefits to be added are in terms of money was actually brought out in Daniel Bernoulli's discussion of the Petersburg problem about 30 years before Bayes. Bayes's argument will still work, however, if the benefits compared or compounded are such as not to interfere with each other; e.g. the benefits to me of two dinners the same night would be far from independent, but those of a letter from you and a bike ride might be nearly so. Ramsey's theory attends to this point, but it is a bit cumbrous.

I'm afraid that more people than you think have the idea that what you say in *Stat. Meth.* is meant for me; and I should guess that about 20 times as many people have read *S.M.* as have tried any of my stuff. If you could find room for a remark that it isn't I should be very grateful. The sixth edition has about half a page blank at the end of the first chapter, which could hold a warning sentence easily!

H. Jeffreys to Fisher: 14 November 1938

. . . By the way who introduced the hypothetical infinite population? Your *Phil. Trans.* paper [*CP* 18] was the first place where I saw it mentioned, but it seems to be implicit in K.P.'s habit of calling chances frequencies in his tables, and it may have a longer history. There is a curious point about Laplace's $(p+1)/(p+q+2)$ rule of succession. It is given in the *Grammar*, and I proved it myself about 1915 for a population of any size — it is exact even for a finite population. I took it for granted that this was as old as the hills, but apparently it was new, at least I can't find it anywhere. Laplace's constant chance at each trial would mean either drawing with replacement or a population so large that the extraction of the sample produces a negligible change in the proportions in it. So perhaps Laplace may be responsible after all!

[P.S.] It also follows, if the sample is all of one type and of number p, the probability that the next s will all be of the type is $(p+1)/(p+s+1)$ whatever the number in the population. There are further extensions.

Fisher to H. Jeffreys: 15 November 1938

. . . I believe the hypothetical infinite population was latent in the earliest notion of sampling error, e.g. in Bernoulli's treatment of the binomial, but it seems to have lain latent for an unduly long time; at least I was conscious of wanting to make it explicit when I coined the slightly provocative phrase you quote. It is just possible, however, that it was previously used by Venn, who, in some passages, was wrestling with the same idea.

No, I did not recognise that the rule of succession was unchanged in form in sampling from a finite population.

H. Jeffreys to Fisher: 9 January 1939

About the question of $ns^2 = \Sigma(x-\bar{x})^2$ or $(n-1)s^2 = \Sigma(x-\bar{x})^2$. In response to wails from my class last term, I have decided that a probability book from me is due, and have started it. I don't want to make unnecessary departures from the usual practice, and in particular I am adopting the practice of using italics for estimates and Greek letters for quantities to be found — it was a matter of random selection after getting fed up with σ and σ' that made me adopt the opposite practice. However I'm bothered about n or $n-1$. The argument for $n-1$ is, I think, that it brings the mode of the z distribution to $z=0$. With n there is a bias, though the formulae are similar enough for one to have to look pretty carefully to see the difference. The z distribution goes over into my theory in an estimation problem: given one set of deviations, what is the probability of a given scatter in a new series of observations or means where there is reason to expect σ to be the same?

On the other hand in the method of least squares what comes in is $\{1+n(\alpha-a)^2/\Sigma c^2\}^{-(n-m+1)}$, and in significance tests for one new parameter the same thing with index $-(n-m+2)$ where in the former m is the total number of parameters found, in the latter the number of old ones; c is a residual. So it is really $\Sigma c^2/n$ that comes in, the m appearing only in the index; and taking $\Sigma c^2/(n-m)$ will complicate the algebra badly. Also it's the sort of algebra that I usually get wrong.

So it rather looks as if I shall have to stick to n. . . .

Fisher to H. Jeffreys: 12 January 1939

I always think that the best interests of mathematical notation are served by each man using the convention which suits him best. Consequently, I am not inclined to argue about using n or $n-m$. I don't quite follow what your letter says about least squares. However, since I always find it convenient to develop the theory in terms of undivided sums, e.g. in the coefficients and right-hand sides of the equations for regression coefficients and in the formulae for obtaining the sums of squares of residuals, as I see it the only question arises when one wants a mean square derived from this sum of squares, and then I am glad you agree that the old procedure of dividing by $n-m$ is the one to use. Of course, there is no great harm in doing what the biometricians do, and dividing all sums of squares by n automatically at any other point in the work, for such divisors cancel out in estimating the regressions. In fact, I regard regression work from another point of view as a good example of ancillary information, in that the precision of the regression does not really depend on the number in the sample, but only on the sum of squares of the independent variate, or, in general, on the dispersion sums of squares and products of the *set* of independent variates. Since the actual values of these are provided by the sample, there is no need to consider estimation in respect of the variances and covariances of these independent variates. In fact the whole work is completely independent of how they may be distributed in the population sampled. . . .

H. Jeffreys to Fisher: 5 August 1939

Bartlett has just sent me another paper on the Behrens–Fisher test, and I see you are having another go too [*CP* 164]. I sent Yates's paper on but have heard nothing further.[1] I have never seen the full form of your function and should be interested to know whether it is the same as my methods give for the *estimation* problem. The significance one is in a *Proc. Roy. Soc.* paper on the comparison of means.

I should state the problem thus. x and y are two unknown location parameters, σ and τ two unknown standard errors, θ and θ' two sets of observations intended to estimate x and y. What, given θ and θ', is the probability distribution of $y-x$? We have (the four unknowns being supposed irrelevant)

$$P(\mathrm{d}x, \mathrm{d}y, \mathrm{d}\theta, \mathrm{d}\tau | H) \propto \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}\theta\, \mathrm{d}\tau\, /\, \sigma\tau$$

$$P(\theta, \theta' | x, y, \theta, \tau, H)$$

$$\propto \sigma^{-m}\tau^{-n} \exp\left(-m\{(x-\bar{x})^2 + s^2\}/2\sigma^2 - n\{(y-\bar{y})^2 + t^2\}/2\tau^2\right)$$

$$\therefore P(\mathrm{d}x, \mathrm{d}y, \mathrm{d}\sigma, \mathrm{d}\tau \mid \theta, \theta', H) \propto \sigma^{-m-1}\, \tau^{-n-1} \exp(\underline{\quad\quad})\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}\sigma\, \mathrm{d}\tau.$$

Integrating with respect to σ and τ

$$P(\mathrm{d}x, \mathrm{d}y \mid \theta, \theta', H) \propto \{1 + (x-\bar{x})^2/s^2\}^{-\frac{1}{2}m}\, \{1 + (y-\bar{y})^2/t^2\}^{-\frac{1}{2}n}\, \mathrm{d}x\, \mathrm{d}y$$

and then putting $y = x + z$ and integrating for x

$$P(\mathrm{d}z \mid \theta, \theta', H) \propto \mathrm{d}z \int_{-\infty}^{\infty} \{1 + (x-\bar{x})^2/s^2\}^{-\frac{1}{2}m}\, \{1 + (x+z-\bar{y})^2/t^2\}^{-\frac{1}{2}n}\, \mathrm{d}x$$

and it only remains to do the arithmetic. At all events it is short and no information has gone down the drain; and it is perfectly clear what has happened to σ and τ.

This is what I call an estimation, not a significance, problem. The distinction is in the form of the question, but also affects the answer. Suppose you get an estimated difference z within the acceptance limit; do you say that the true value is as likely to be $2z$ as 0, or that it is more likely to be 0? If the former, it is what I call an estimation problem; if the latter, one of significance. It seems to me that most direct comparisons in agricultural experiments are estimation problems, but if you are testing whether a Mendelian ratio agrees with 3:1 it is a question of significance.

It seems to me that in your original fiducial argument you were speaking of the probability of the true value, given the data, in precisely the same sense as I should, and you and 'Student' both make the point clearer in various places by emphasizing the uniqueness of the sample, thus saying what I say by using the prior probability that expresses previous ignorance. The trouble is that your language is not adequate to get into symbols what you are trying to say, and consequently you don't state all the steps that I filled in in my paper[2] about 'Student'. . . .

[1] See Fisher's letter of 20 February 1939 (p. 352).
[2] See Jeffreys' letter of 18 May 1937 (p. 162).

Fisher to H. Jeffreys: 8 August 1939

Many thanks for your interesting letter. I think your analysis is identically equivalent to mine and would lead to the numerical values given by Sukhatme. I hope you know his table, as it is a very usefully arranged one. I see, and like, your distinction between *estimation* and *tests of significance*, but am not quite sure of its application in this case. You are probably right that the language used by 'Student' and myself is not adequate to get into symbols what we are trying to say; but, of course, the effort to fill in all the steps is full of the danger of saying something quite different.

I am beginning to realise that we know much more about the methods we practise in reasoning than about the systems of logical postulates necessary to justify these methods. Consequently, such a question — which sounds urgent and important to purely deductive minds — as: 'Is there any new principle or postulate required or assumed in this method?' cuts very little ice with me because I do not know that the principles or postulates required for previous methods have ever been satisfactorily taped. Unless that has been done, and I respect Whitehead and Russell, Keynes, etc. as I respect the leaders of forlorn hopes, the first question asked is really meaningless. After all, may not the recognition of logical rigour be as empirical as the recognition of the three-dimensionality of space? . . .

Fisher to H. Jeffreys: 4 November 1939

I have just re-read your note on the Behrens-Fisher formula, which you sent me some time ago, and I forget whether you were inclined to publish it in the *Annals of Eugenics* or in the *Camb. Phil. Soc. Journal.* It would seem perfectly suitable for either.

I find I can follow your arguments perfectly, and should disagree, if at all, only on terminology, for you use the distinction between tests of significance and estimation differently from the way I do, including in the latter cases in which the answer is not, properly speaking, an estimate. However, apart from this and the propriety of using the *a priori* factor $\frac{1}{2}$ when a precise null hypothesis is specifically in view, I think your paper enables me to appreciate your point of view a great deal better than I have previously done.

I have already printed in the *Annals* rather more than I should have liked on what is really a technical point in theoretical statistics, with no genetic or human applications immediately in view; but, if you prefer the *Annals* to the *Cambridge Philosophical Journal*, I should be glad to put it in the first number of Vol. X.

H. Jeffreys to Fisher: 6 November 1939

Glad to see 'The Galton Laboratory' is still — or again — a recognized address.

I should prefer the paper to go to the *Annals*. . . .

I am not sure what cases you have in mind when you say that I include in estimation problems cases where the answer is not an estimate. In those I had in mind it seemed to me that there was always an estimate not very far away. E.g. suppose you have m sets of n observations; one standard error is estimated from the means and another from the scatter within the groups. If the former is too large it suggests a further variation, possibly normal, affecting the whole of a group, and the comparison leads to an estimate of this. But this is a typical z problem. I have extended this to unequal groups, as this trouble is continually turning up in astronomy and geophysics; there is no neat answer, but an approximate one can be got. . . .

Fisher to H. Jeffreys: 8 November 1939

Thanks for your letter. I will put your paper into *Annals*, Vol. X, No. 1.[1] . . .

[1] Jeffreys, H (1940). Note on the Behrens–Fisher formula. *Ann. Eugen.* 10, 48–51.

H. Jeffreys to Fisher: 29 March 1942

Do you know anything about the following problem? It turns up pretty often, but I have just had it again in a rediscussion of H.S. Jones's determination of the mass of the moon. We have two sets of estimates of x, with different standard errors, estimated from n_1 and n_2 observations. There is no doubt about whether the two estimates of x refer to the same thing, so it is an estimation problem. The question is, how to allow for the difference of the estimates in finding the uncertainty of the combined estimate of x? The exact solution is

$$P(\mathrm{d}x|\theta, H) \propto \mathrm{d}x\,\{1 + (x-\bar{x}_1)^2/s_1^2\}^{-\frac{1}{2}n_1}\,\{1 + (x-\bar{x}_2)^2/s_2^2\}^{-\frac{1}{2}n_2}$$

so that it is the Behrens–Fisher rule with the two x's the same and no integration therefore to get a distribution for the probability of their difference. But can we get a useful approximation to this anyhow? The point

is that in my problem $\bar{x}_1$ and $\bar{x}_2$ are based on $n_1 = n_2 = 13$ but differ by a shade under twice the standard error of their difference. This is really sure evidence that s_1 and s_2 are estimated a little too low and we could get a better idea of the uncertainty of x by allowing for the other degree of freedom expressed by $\bar{x}_2 - \bar{x}_1$. If only the standard errors in the two sets were equal it would be easy, but they aren't.

I see you have a new paper on this sort of thing but it can't be borrowed yet and I don't know exactly what you've done.

Fisher to H. Jeffreys: 30 March 1942

You are quite right that the problem you mention is very near to that of Behrens, but owing to the additional degree of freedom interpreted now as additional evidence on the values of s_1 and s_2, the significance test tabulated for Behrens' test is not exactly applicable. The analysis is, however, very similar in form, and I was tempted to try to tabulate values for your problem, and only deterred by there being one more (4 instead of 3) parameter with which the table must be entered. Tabulation would therefore be rather extravagantly elaborate, and the use of effective approximations based on the asymptotic approach of 'Student's' to the Normal distribution is all the more important.

I have discussed this approximation rather fully in the enclosed paper [*CP* 181]. . . .

The fundamental relations used are set out in Section 2, and the asymptotic approach in Section 3. Your case differs from Behrens in that, though both depend on a well determined line in the plane of t_1 and t_2, Behrens is concerned with the total frequency integrated on one side of the line, whereas you have to deal with the section of the frequency surface cut by the line itself. The maximal likelihood solution is the mode of this section, and to assign fiducial limits to this estimate one is concerned to determine the points on the lines which cut off $\frac{1}{2}$%, or $2\frac{1}{2}$%, or whatever you want of the sectional area at each tail, i.e. at p. 152 you would put $y = d$ in the expression for t_1 and t_2 and integrate for x only. I imagine the analysis will be about as difficult as what I have done on this and the next page, but you may not need so many approximative terms as I have introduced. At least, using definite values from the start, you are in a good position to see where to stop, whereas with algebraic investigation the relative importance of different terms is different in different parts of the table.

In the new edition of *Statistical Tables* also I am putting table 6 of this paper, together with Sukhatme's tables and an illustration of the sort of application I should think appropriate in physical or astronomical work. I should be very glad to hear what you think of these two parallel attempts, especially from the point of view of logical exposition, since such a disturbance has been raised about this aspect of the problem.

H. Jeffreys to Fisher: 9 *April* 1942

Thanks for your letter. The extra parameter finishes the thing off! . . .

I haven't anything to say about Sukhatme beyond what I said in my *Ann. Eugen*. paper — that it's right as a method of estimation but it isn't what I should call a significance test.

By the way you are inclined to blame omissions in the older methods on the teaching of inverse probability. I should rather blame it on the fact that it took a long time for people to see that getting the right answer depends on stating the question properly. With inverse probability it can be stated rightly or wrongly, but at least it has to be stated. Without it there is a continual risk of muddle through people not seeing what the problem is. As a matter of fact in 1921 or so I did the 'Student' problem using $P(dh \mid H) \propto dh$, but wasn't happy about it and didn't see how to put it right till *Scientific Inference*. But the difference between n and $n+1$ in 'Student's' formula is far less than the difference between you and Bartlett!

The awful problem in physics etc. is to get people interested at all. They seem to like multiplying by 0.6745 if they can be bothered to work out an uncertainty at all, and any attempt to tidy up the theory rather annoys them because it restricts their liberty of guessing. So for your tables I should say that it is no earthly use putting in physical applications that involve any appreciable amount of arithmetic. . . .

Click here for next section