*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 $P(\mathrm{d}h \mid H) \propto \mathrm{d}h$, 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. . . .

---

*C. Jordan to Fisher:* 14 *May* 1934

I thank you very much for your letter and the highly interesting paper you kindly sent me. I know that you establish your theory of maximum of likelihood independently of the Theory of Probability. I have shown that your method is equivalent to the following; an approximation of $y_i$ by $f(x_i)$ is considered as being so much better as the geometrical mean of the geometrical deviations $y_i/f(x_i)$ is less (for $i = 0, 1, 2, \ldots n$).

This is quite different from the older methods of approximation, in which the minimum of a function of arithmetical deviations (standard deviation) is required. The geometrical mean and deviations seem to be according to Weber's and Fechner's law of psycho-physics much more important to us than the arithmetical mean and deviations.

Your method, applied to Laplace's and Poisson's formula of repeated trials, conducts, as does the method of moments (but not the other methods) to the usual results.

As you reject Bayes' theorem, it is all the more remarkable that the method of approximation according to the principle of maximum of likelihood is a necessary consequence of Bayes' theorem applied to Bernoulli's rigorous formula of repeated trials.



This is interesting, since there are many (I am among them) who think that it is impossible to form a rule for choosing between two hypotheses without admitting some postulate, and who think that the postulate of equal probability of the different possible hypotheses is at least as good as any other; for them Weber and Fechner's law and Bayes' theorem are the justifications of your method which show that it is preferable to any other method of approximation. . . .

*Fisher to C. Jordan:* 18 *May* 1934

. . . The parallelisms you mentioned are exceedingly interesting. I cannot help feeling, however, as indeed you recognise, that the logical basis of the use of likelihood must be really independent of Weber and Fechner's Law; i.e. of the mechanism by which differences in the intensities of a sensation are appreciated, especially since this law can only hold over a definite range of the intensity of the stimulus.

The connection with Bayes' Theorem, or rather with the postulate, which Bayes proposed, after proving his theorem quite rigorously on the data of the problem, has a more intricate history. Supposing that the postulate were true of a particular unknown parameter $\theta$, i.e. it is known with certainty *a priori* that the probability that $\theta$ lies in the range $\mathrm{d}\theta$ is proportional to $\mathrm{d}\theta$ for all values of $\theta$, then Bayes' method gives us an inverse probability distribution of $\theta$ having, if $\theta$ is a continuous variable, a determinate form, and therefore determinate values for its mean or mode or median or any other property of the distribution in which we might be interested. If I believed in this inverse probability distribution, I should be most interested in those of its properties, such as the median, which are invariant at least for a wide class of functional transformations of the variable $\theta$. I should attach less importance to features such as the mean and mode, which may be brought to any value at will, by an appropriate functional transformation. The method of inverse probability, therefore, based on Bayes' postulate, does not point with any special emphasis to the mode of this distribution which, as is well-known, is in fact the solution of the equations of maximum likelihood. In arriving at the mode by this method two arbitrary steps have been taken, neither of them, in my opinion, justifiable, which in effect cancel each other out: (i) the choice of the particular parametric function $\theta$, to which to apply Bayes' postulate, and (ii) the choice of this particular function in relation to which to determine the mode of the inverse distribution. If the same function is used for both processes, the maximum likelihood value is reached and this is in fact invariant for all functional transformations, and though Gauss justified his choice of the maximum likelihood by the inverse probability argument, I feel very sure that he would not have in fact adopted it if it had not been invariant.

I think it is well worth while to get to the bottom of the confused history of this intricate subject, since at the present time it seems possible at last to put at least some useful processes of inductive reasoning on an entirely rigorous basis.

---

*Fisher to M.G. Kendall:* *2 November* 1942

I was much interested in your paper 'On the future of statistics',[1] and should have made a point of being present if any notice had reached me.

On p. 73 you ascribe to Neyman and Pearson a certain advance in the theory of inference. I could refer you to a good many things I had published on the subject before Neyman and Pearson knew anything of what I was doing; but in view of what you say I really think you ought to read a short note published in 1930, [*Proc.*] *Camb. Phil. Soc.* Vol. 26: 528–535, which I wrote under the title 'Inverse probability' [*CP* 84].

P.S. When you have read the reference I think you will understand why I should look to you positively to combat the fable of Neyman and Pearson introducing fiducial inference, for others beside yourself have been deceived by the attitude they have taken up.

[1] *J.R. Statistic. Soc.* **105**, 69–80 (1942).

*M.G. Kendall to Fisher:* *7 November* 1942

Thank you very much for your letter of 2nd November. I am glad you found some interest in my speculations and am sorry you were unable to be present. I assumed at the time, of course, that you would receive a notice on a card in the usual way, but apparently yours must have gone astray. Had I thought there was any danger of such a thing I would have made a point of letting you know specially, as I am sure the discussion would have benefitted from a contribution by you.

As regards inference I am sorry if you feel I have not given credit in the right quarter, but I am sure you know me well enough to accept my assurance that anything of the kind was quite unintentional. The question of confidence intervals versus fiducial probability still puzzles me a lot. Like several other people, I had supposed that they were rather different ways of looking at the same thing; but from the argument between yourself and Bartlett and the recent paper by Neyman in *Biometrika*[1] it appears that this is not so, since the two methods can give different answers to the same question. In more peaceful times I would have tried to get over to Harpenden to discuss it with you, but what used to be half-an-hour's trip in the car has now become a full day's journey.

An opportunity to set out the whole position fully and fairly will, I think, present itself when I get down to writing the second volume of my book on the Advanced Theory of Statistics. The first volume is now in the galley proof stage and ought to appear some time next spring; but it does not deal with inference systematically, although I have introduced likelihood on more or less intuitive grounds. The second volume will take about another two years to write, even if the war comes to an unexpected end, and I very much hope that during that time we may have an opportunity of a talk about the whole position. It is part of my programme to give an exposition of fiducial inference, but there are several points on which I should like to be sure that I have grasped exactly what you mean. I suppose you couldn't spare the time to lunch with me the next time you are in town? I can usually be found at the Chamber of Shipping, Avenue 5381.

P.S. You will, of course, receive a copy of my book from the publishers.

[1] Neyman, J. (1941). Fiducial argument and the theory of confidence intervals. *Biometrika* **32**, 128–50.

*Fisher to M.G. Kendall:* *9 November* 1942

Thanks for your letter. I am sure there are plenty of differences, not only in exposition, but also, very probably, in results, between my approaches and Neyman's. However, the particular innovation of reasoning which you specified to the Statistical Society in the words[1]:

> 'The essential concept of the theory of confidence intervals, if I have understood it correctly, is the replacement of these specified limits by . . . functions of the sample numbers.'

is to be found quite explicitly in the paper to which I referred you. I may say it took a lot of private coaching by me to get this simple modification into Neyman's head. When he had grasped it, he proceeded, with Pearson's help, to expound it to the world with the minimum of reference to its origin.

I should love to meet you for lunch some time, and am fairly often in town on A.R.C. business, but their meetings usually make one too frightfully late. Is the Statistical Society likely to make an occasion?

[1] *J.R. Statist. Soc.* **105**, p. 73 (1942).

*M.G. Kendall to Fisher:* *23 October* 1943

The publishers have just sent me a copy of your review of my book[1] in *Nature*, and I should like you to know how much I appreciate some of the kind things you have said and your understanding of the difficulties of writing it.

Perhaps the second volume will be more after your own heart. It has certainly been more interesting to work on, though more difficult to pull into line. I have nearly finished a first draft, although a further six months will be necessary to put it into final shape for the printers.

I had been intending asking your help on volume two in the following respect. In dealing with estimation I decided to have four chapters: Estimation by Likelihood; Estimation by other 'point' methods (e.g. minimum chi-squared, least-squares, the Aitken–Silverstone approach and so forth); Estimation by Confidence Intervals; and Estimation by Fiducial Inference. It is the last two which present the difficulty. Until the controversy over the Behrens–Fisher test I had supposed that confidence intervals and fiducial intervals were essentially two different ways of looking at the same thing. It now appears that this is not so, and to avoid any possibility of mixing up the two I decided, as I say, to have them in separate chapters.

I hardly think it would be proper for me to intrude my own personal views as to the relative merits of the two methods. In a book such as this I am anxious to give an objective presentation. The two chapters, therefore, will attempt to give an account of each theory and to explain the differences between them, but will not reject one in favour of the other.

I think you will agree that it is very important that the presentation of each theory shall be fair and accurate; and I wonder therefore whether you would be so kind as to read the draft of the chapter on fiducial probability and let me know whether you approve? It will not be ready in a readable form until some time after Christmas so there is no hurry. To safeguard myself on the confidence interval side I am asking Egon Pearson to read the chapter on confidence intervals. . . .

[1] *The advanced theory of statistics*, Vol. 1. Griffin, London (1943). For Fisher's review, see *Nature* **152**, 431.

## *Fisher to M.G. Kendall:* 3(?) *November* 1943

Thanks for your letter. I shall be certainly be glad to do all I can to help, but you have, even in your letter to me, already adopted part of the language and terminology which Pearson rather maliciously as I thought introduced (after ignoring my work for fifteen years) with a view to misleading the student as to its scope and nature. The so-called 'point' or 'single-value' methods of estimation constitute the whole problem or subject of estimation in the sense in which this phrase was used for at least fifteen years. When later I introduced the fiducial argument and showed how fiducial limits could be calculated, this did not in my opinion constitute a new method of estimation but rather a new method of constructing probability statements valid in respect of the unknown parameter, obviously quite a closely related topic, but one applicable to a smaller range of cases.

The criteria of estimation originally developed were based in respect of efficiency on the variance of the sampling estimate, a fact which alone shows how misleading is the phrase, 'single-value' or 'point' estimation, applied to such methods. In respect of sufficiency the exact form of the whole sampling distribution is taken into account or, looked at otherwise, the entire course of the likelihood function.

For several years it is certain that Neyman thought, in speaking of 'confidence-intervals', that he was only systematising and developing a a new



exposition appropriate to the fiducial argument I had put forward and pressed upon his notice. . . .

The diversity of opinion as to Behrens' problem arises, I think, by an historical accident. In our statements of probability, e.g. tests of significance, neither 'Student' nor I took the problem to relate to a population of samples all of the same size and drawn from an identical population. In 'Student's' problem we were both thinking of a population of samples of the same size drawn from populations having their variances fiducially distributed, i.e. that unique distribution of populations for which alone the observed estimate $s^2$ would be a typical estimate. It so happens, however, that when there is only one unknown parameter, and in a few other cases of special simplicity, the probability which defines the level of significance can be reinterpreted as a probability related to a population of samples drawn from the same population. So to interpret them is to my mind always to wander from the point, but Neyman and Pearson have made this convention the headcornerstone of their exposition and, as the Behrens' case shows, are much aggrieved when what is frankly the only possible solution of a practical problem does not conform with their convention. This however will generally be the case whenever tests have to be made in situations involving multiple estimation. It is indeed surprising that Neyman and his followers, who so often discuss a comparison of two samples, should be unable to provide any solution to this rather fundamental and primary problem. I presume, of course, that the method, following Bartlett, of choosing one out of $n!$ equally eligible tests of significance and using its results without reference to that of any of the others, is not to be regarded as a solution of the problem. Moreover, the hope, which I suppose Bartlett originally entertained, that there was an appropriate test which would reject a fixed proportion of samples for all values of the variable parameters when the samples are taken to be drawn from a fixed population, seems now to have been finally abandoned.

One fundamental objection to basing the theory on samples of a fixed size from a fixed population is the obstacle it introduces to the rational use of ancillary information. Properly speaking, the sample-number or set of such numbers used to specify the sample are ancillary statistics, and in particular cases are found to be supplanted by more appropriate ancillary statistics, though of course in other cases such statistics are additional to the sample-number. When, in such cases, estimation is exhaustive, the likelihood function is completely specified by the set of ancillaries, and it can be shown that all other statistical calculations add precisely no information to that already in hand; it seems absurd to pretend that there is a fundamental distinction between such cases and those which I originally called 'sufficient', which differ from them only in that in these the sample-number alone supplies all the ancillary information required. Yet in the method of exposition adopted by Neyman and Pearson these cases are separated from the first, and indeed the use of ancillary information seems to be almost ignored.

### *M.G. Kendall to Fisher: 7 November* 1943

Thank you for your very interesting letter. I will send a chapter along when it is ready and hope you will be able to approve. In the meantime, a few comments on your letter:

I haven't used the 'point-estimation' phraseology at all up to the present. It must have slipped into my last letter because I had just been reading Wilks' latest effort on Mathematical Statistics (have you seen it? — I wasn't very impressed). As I understand it you conceive of the estimation problem as that of determining a value surrounded, so to speak, by an unavoidable band of doubt; whereas in Neyman's treatment (as he now states it) the interval is everything and there is no pretension of locating the estimate at one point rather than another.

I don't know what you thought of Jeffreys' comments on the matter in the *Ann. Eugen. 10*, 48,[1] but he seemed to me to have made a point when he compared your attitude with his own towards posterior probability. Yates also speaks of the fiducial distribution as a sort of distribution of doubt, with a mode at the estimate and tailing off on either side.

Would you agree if it were put this way? If one is not prepared to make any additional assumptions or postulates, or to restrict the nature of the inference, one can get no further than Neyman and the Behrens' problem is insoluble. It becomes so if one is prepared to [do] one of three things: (a) follow Jeffreys and use probability as a measure of rational doubt, not necessarily related to frequency; (b) restrict the inference to the family of parents for which the sample variance is distributed in the specified Type III form: (c) adopt a new postulate to the effect that differentials like ds can be replaced by those of type-d$\sigma$. Probably (b) and (c) are the same thing in the ultimate analysis and no doubt Jeffreys would maintain that they are both equivalent to (a) — though I am always nervous about saying what he might think because I never seem to express his views in the way he would like.

[1] See Fisher's letter of 8 November 1939 to Jeffreys (p. 176).

### *Fisher to M.G. Kendall: 10 November* 1943

Thanks for your note. I suppose that all science, including all but the most highly purified mathematics, requires some such postulate as that observations are not generally made by especially privileged or exceptional observers. For most human purposes, though not for cosmology, something of the kind is needed merely on the ground that what is going to be called science is to be public knowledge, true for all (or nearly all) mankind, whereas the observations of the exceptionally privileged observer are by definition available and repeatable only by an equally privileged one. In cosmology the argument seems to take a different form but to be equally necessary, as is well illustrated by Milne's statement of the same principle.

The form the postulate takes when faced with what 'Student' called a 'unique sample' is that the only basis upon which a unique body of information subsumed in such an estimate as $s^2$ can be utilised is that $s^2$ is typical or representative, or not selected, biassed or sophisticated. On the other hand, for refined purposes, we shall never want to assume that $s^2$ is equal to the true unknown variance of the population sampled.

The known sampling distribution for $s^2$, given $\sigma^2$, then provides us with a well-defined random variable $\sigma^2$ such that for the population of parent populations having such a random variance, $s^2$ is a typical or representative estimate. This, I imagine, was clearly implicit in 'Student's' argument in 1908 and made quite explicit in my paper introducing Fiducial Limits in the Cambridge Philosophical Society's proceedings [*CP* 84].

I doubt if axiomatics go to the root of the difference of opinion, for any rational approach to the interpretation of data must contain some axiom justifying the propriety of using the data at all, or basing any inferences upon them. So far as I can see, the difference is due wholly to Neyman having seized upon an unessential feature of the simpler tests of significance in wide use when he first became acquainted with the subject, namely that the level of significance was in fact, in these simpler cases, the proportion of samples obtained by repeated sampling of a fixed population which would be rejected by the proposed test. This, I think, only *happens* to be true; the tests are good not because it is true but because of the different proposition that the proportion of the population of samples, of which that presented is a representative, which are rejected by the test is in fact the fraction specifying the level of significance.

As to axiomatics, I have quite an open mind. I do not believe that a complete system of axioms required for arithmetic or analysis has ever been successfully laid down. I am sure this would have been done if it were easy: so I suppose it is difficult, and that the satisfactory and flourishing development of a subject need not wait upon the precise definition, if this is ever achieved, of the system of axioms from which the subject might (as an academic exercise) later be developed.

---

### *N. Keyfitz to Fisher: 18 November* 1955

I was delighted to see your article [*CP* 261] in the number of the *JRSS* which has just come in. It makes clear some very important issues about which I have been puzzled; several of my colleagues here join me in thanking you.

If we understand you, the difference between the situation of an acceptance sampler and that of an experimenter is the difference between facing a problem of human devising and facing a problem presented by nature. The acceptance sampler who has to discriminate between two distributions may perhaps be compared with a student of probability discriminating between two urns in a textbook problem exemplifying Bayes' Theorem. To make the investigation of nature analogous to sampling a controlled production process apparently is only useful if the capacity of nature for offering surprises is of the same order as the production process. You presumably fear that to promote acceptance methods in science does more harm than merely distorting the meaning of some words like 'decision' and producing irrelevant mathematical embroidery; it helps hide from the investigator the tentative process by which alone he can uncover new facets of the real world. The ideological implications of a mechanical decision theory are terrifying if one thinks about them along the line to which your article points. . . .

### *Fisher to N. Keyfitz:* 21 *November* 1955

I was very glad to get your letter of November 18th, and I may say that from other sources I learn that the article has led to a healthy reconsideration of what we are doing in some fields, and I hope it will lead to a recognition of when we are and when we are not attempting to add to natural knowledge.

I think it will be common ground in the technological fields generally that the technologist is constantly concerned with fact finding, or with seeking verification of provisional hypothetical suppositions of the scientific type. In quality control, for example, I imagine that the high level work is all of this kind, involving the whole range of concepts of experimental design, and that only the warning bells and red lights can usefully be mechanized in an automatic acceptance procedure.

What is particularly troublesome is that Neyman, in importing from Eastern Europe his misconceptions as to the nature of scientific research, should have chosen so ubiquitous a scientific tool as the test of significance as the subject on which to fasten ideas relevant only to the acceptance procedure. A typical test of significance is based on a probability statement derived from the hypothesis to be tested, and therefore existing only in the hypothetical world created by this hypothesis. Typically it leads to no probability statement in the real world, but to a change in the investigator's attitude towards the hypothesis under consideration, for which if we choose to use the word 'rejection' we must remember that the rejection is only provisional, and that our hypothetical calculations have shown that there would be a finite probability of our obtaining the observed level of significance even were the hypothesis true. In fact future observations may later demonstrate, with all the force of which scientific evidence is capable, that it is really true, and the the provisional rejection was due to an exceptional coincidence. Of course we may on occasions take irreversible action on the strength of such evidence, but such a possibility does not at all affect the logical status of the evidence upon which it is based. . . .



### *Fisher to T. Koopmans*: 29 *November* 1935

. . . As I expect you noticed, in my paper, 'Two new properties' [*CP* 108], I dissent from the view expressed by Professor Pearson and Dr. Neyman that the results they have arrived at cannot be obtained by the theory of estimation.[1] Indeed, I think their remark was only made in self-defence from the obvious criticism that the only result of practical importance to which their methods had led was a restatement of an easily demonstrable property of sufficient statistics.

As I did not introduce the term 'confidence intervals' I cannot object to its use in any circumstances for which it was intended. By themselves they seem to be only another way of saying that, by a certain test of significance, some kinds of hypothetical possibilities are to be rejected, while others are not. This is not what struck me as a novelty when I first put forward the notion of fiducial probability, namely, that when properly defined, rigorously exact probability statements could be made about the values of unknown parameters, the whole set of such statements providing the fiducial frequency distribution of these parameters. If for this purpose non-sufficient estimates are employed, the statements of fiducial probability and the frequency distribution which comprehends them will be as definitely erroneous as if we had arbitrarily rejected or ignored a portion of the available data and based our conclusions upon the selected remainder.

With respect to the general form of the parent distributions, when sufficient estimation is possible you may be interested to see that I arrive at conclusions very like those which you give, in the section of the paper on 'Two new properties' preceding that to which you call attention, the section in which I develop the general sampling distribution of sufficient statistics for simple estimation. I suppose, though I have not myself verified it, that the simultaneous sampling distribution of a set of statistics sufficient for the estimation simultaneously of several parameters could be developed in the same way.

[1] Koopmans had written asking (i) whether there was a 'real difference between the approach of Professor Pearson and Dr. Neyman in their papers on hypothesis testing and that which you call "theory of estimation" ' and (ii) what were the grounds on which Fisher, in his contribution to the discussion of Neyman's 1934 paper on the representative method (*J.R. Statist. Soc.* 97, 558–606), 'deprecated the use of "confidence intervals" based on non-sufficient statistics'.

### *Fisher to H.E. Kyburg:* 12 *January* 1962

I have just read your letter of Oct. 17, which must have travelled a long way since it left your hand; moreover it is so kindly expressed that you must have been puzzled at not receiving a prompt reply.

I only realised the backwardness of many departments in U.S. purporting to teach mathematical statistics, when I put in a term at Michigan State (I fancy it was in the Fall of 1956 [1957], at least the first Sputnik went up while I was there). I was in fact very deeply shocked not only by the poor material being taught to the students, but by the spirit of ideological propaganda, in which other views were 'bawled down'. That seemed to me to have no place in a centre of learning. . . .

However, just in this last year, those in the States who thought there was some sense in what I had been saying, have been reinforced by two vigorous voices from across the northern frontier: Donald Fraser (Toronto), whom I have not yet met, and D.A. Sprott from a rather unknown University (Waterloo) whom I met last May in George Barnard's department at Imperial College. So the moulding of mass opinion has not been fully successful. Sprott thought that *no one* in U.S. would agree with him. I think he knows better now. . . .

*Fisher to H.E. Kyburg: 22 February* 1962

I believe in answering your most flattering letter I may have misunderstood you on one point. I supposed that your book was already published, and intended to procure a copy.

It now seems possible that this is not so, and that you were offering me the favour of reading typescript or proof prior to publication. I should at once accept this kindly offer; for though I am blind as any bat, I can still do some reading, and I am indeed immensely curious that an approach through Symbolic Logic should have so closely concurred with my own approach through experimentation.

I should hate you to think that I had ignored so friendly a gesture.

I expect you will have heard from Rao in Calcutta that he wants you to talk, if you can bear it, at a symposium of the International Statistical Institute at Ottawa in the Summer of 1963.

With luck I hope to be there.

[P.S.] I am at work on the rather difficult job of specifying when, and when not, the data supply the materials for fiducial probability inferences.

*Fisher to H.E. Kyburg: 14 May* 1962

After a long while I have now succeeded in obtaining your book on Rational Belief.[1] So far it seems to be as good as I had hoped, which would indeed be high praise.

As you know I am inclined to stratify Keynes' wide notion of degree of rational belief, according to the nature of the evidence, into Mathematical Probability in the sense of the 17th century mathematicians which does not include the whole of uncertainty as Voltaire and Montesquieu assumed, and a number of humbler states of uncertainty including the Likelihood Statements, almost always available in scientific work, and which have recently been rediscovered by Birnbaum, whom perhaps you know. I forget if I ever sent you a puzzling paper called the Underworld [*CP* 267].

Anyway let me know if ever you want any such thing of mine. I wonder if the Underworld could be translated into your logical symbolism. . . .

[1] *Probability and the logic of rational belief.* Wesleyan University Press (1961).



*J. Neyman to Fisher: 9 February* 1932

. . . I do not know whether you remember what I said, when being in Harpenden in January 1931, about our attempts to build the theory of 'the best tests'. Now it is done more or less. It follows from the theory that the 'Students' test, the '*t*-test' for two samples and some tests arising from the analysis of variance are 'the best tests', that is to say, that they guarantee the minimum frequency of errors both in rejecting a true hypothesis and in accepting a false one. Certainly they must be properly applied. You will see that these results are in splendid disagreement with the last articles of Professor Pearson in *Biometrika*.

Presently Dr. Pearson is putting all the results in order. They will form a paper of considerable size. We would like very much to have them published in the *Philosophical Transactions*, but we do not know whether anybody will be willing to examine a large paper and eventually present it for being printed. The paper contains much of mathematics and not all the statisticians will like it just because of this circumstance.

We think that the most proper critic are you, but we don't know whether you will be inclined to spend your time to read the paper. . . .

*Fisher to J. Neyman: 12 February* 1932

. . . I should be very much interested to see your paper on 'the best tests', as the whole question of tests of significance seems to me to be of immense philosophical importance, and the work you showed me was surely of great promise. It is quite probable that if the work is submitted to the Royal Society, I might be asked to act as referee, and in that case I certainly shall not refuse.

*J. Neyman to Fisher: 16 October* 1932

Reading the paper by A.T. McKay: 'Distribution of the coefficient of variation and the extended *t* distribution' in the *J.R.S.S.*, Vol. XCV, Part IV, p. 695, I have been alarmed by the information that you have been doing something with this distribution and that the results are published in the B.A.M.T.[1]

As a matter of fact the said distribution is being tabled in this laboratory and the work is already fairly advanced. Unfortunately our financial conditions do not allow us to purchase the B.A.M.T., at least in the nearest future, so I beg you to be kind enough to inform me what sort of tables concerning the distribution of $t$ do they contain.

E.S.P. writes that you have recommended our paper for publication. Although it may be considered ridiculous to thank a judge, I have intense feeling of gratefulness, which I hope you will kindly accept. . . .

[1] British Association *Mathematical Tables.*

### *Fisher to J. Neyman:* 24 *October* 1932

It is a great pleasure to hear from you again. About five years ago in order to encourage Airey to calculate the successive integrals of the Probability integral, I drew out some notes on the theory of the functions concerned, their connection with the Hermite functions, and their statistical use. My notes came with Airey's table into the hands of the British Association Committee, from whom, on account of their incompetence and extravagance I was later obliged to dissociate myself, and they published my notes [*CP* 91] in the Introduction without submitting a proof to me, or giving me an opportunity of obtaining offprints. E.S. Pearson is a member of the Committee and might have been able to tell you what the book of tables contains.

My note gives the theory only, giving the solution in terms of the *Hh* functions, which are pretty well tabulated by Airey. Tables of chosen percentiles for varying $n$ and $\tau$ would of course be much more useful in application. I am surprised and interested to hear that you have been working on the same distribution; but before assuming this you ought to be sure that it is the same.

I was a good deal interested in your paper,[1] but I ought not to have passed your statement that a distribution is completely determined by its moments. Do you not know Borel's distribution with all its moments zero?

$$e^{-4\sqrt{x}} \sin {}^4\sqrt{x}\, dx.$$

You can always add a bit of Borel to any distribution without altering the moments, and if your distribution has tails less steep than $e^{-4\sqrt{x}}$, you need get no negative frequencies. . . .

[1] Eventually published as Neyman, J. and Pearson, E.S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans.* A **231**, 289–337.

### *J. Neyman to Fisher:* 28(?) *October* 1932

Many thanks for your kind letter. I am glad to hear that the work of my students is not lost. On the other hand I am rather troubled by your criticism, which seems to be quite correct. The origin of our mistake is the same statement in P. Lévy's book, concerning the continuous distributions. I know of course that the set of moments does not necessarily determine the frequency distribution and that there are means to decide whether it does or not. /See f.i. the papers of H. Hamburger 'Ueber die Erweiterung des Stieltjesschen Momentenproblems', *Mathematische Annalen* Bd. 81, 82./



Unfortunately the criteria are not easy to apply, so — being very much occupied with the new notions of best critical region etc. — we believed too much in what said Lévy.

At present I see that a restriction is necessary.

Namely our theory holds good in cases when the set of moments of $\phi$ — if you remember the notation — leads to a unique frequency function, and this question may be solved by means of the criteria, given by H. Hamburger.

As a matter of fact, I think that in all practical examples we have considered, the theory holds.

Of course there is lot of questions to be worked out. There are too many restrictions.

I am often thinking that it would be very useful for me to work in contact with you. Unfortunately this requires considerable amount of money — without speaking of your consent — of course.

### *Fisher to J. Neyman:* 1 *November* 1932

You may be sure of my consent, if ever it becomes possible for you to work with me. The difficulty you speak of arises I believe solely through working with the moments instead of with the characteristic function, in the Taylor expansion of which, if it has one, the moments appear. I do not think it affects your work on the best critical region.

### *J. Neyman to Fisher:* 3 *December* 1932

A letter more! I am afraid you have already sufficient of correspondence with Poland. However the question seems to be important. We are considering with E.S. Pearson what sort of reference it will be necessary to give in the paper about your showing us the inaccuracy with the moments etc.[1] Can it be: Formerly we did not see that this restriction* is necessary and we are indebted to Dr. R.A. Fisher, who kindly called our attention to the fact that the set of moments does not necessarily determine the corresponding frequency function.

*Namely: 'in cases when the set of moments . . . etc.'

Will it do?

I hope to be in England just after the 20th of December. But the proofs may come any time. So perhaps you will be kind to answer — at least if you do not like the proposed text. . . .

[1] See Neyman, J. and Pearson, E.S. (1933) On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans.* A **231**, p. 315.

*Fisher to J. Neyman: 7 December* 1932

Do not think of thanking me especially for the little point I made about the moments. It really does not affect your paper and I only mentioned it casually because I was writing to you. You and Dr. Pearson are not indebted to me for anything more than, as it were, the correction of a verbal slip.

All the same, I like hearing from Poland. Best wishes for a Merry Christmas.

*J. Neyman to Fisher: 13 June* 1933

Dr. Pearson writes me that soon you will be Galton Professor at the University College, London. Very probably this means a general reorganization of the Department of Applied Statistics and possibly new people will be needed. I know that there are many statisticians in England and that many of them would be willing to work under you. But improbable things do happen sometimes and you may have a vacant position in your Laboratory. In that case please consider whether I can be of any use.

*Fisher to J. Neyman: 17 June* 1933

Many thanks for your letter of congratulation. You will be interested to hear that the Dept. of Statistics has now been separated officially from the Galton Laboratory. I think Egon Pearson is designated as Reader in Statistics. This arrangement will be much laughed at, but it will be rather a poor joke, I fancy, for both Pearson and myself. I think, however, we will make the best of it. I shall not lecture on Statistics, but probably on 'the Logic of Experimentation', so that my lectures will not be troubled by students who cannot see through a wire fence. I wish I had a fine place for you, but it will be long before my new department can be given any sort of unity and coherence, and you will be Head of a Faculty before I shall be able to get much done. If in England do not fail to see me at University College.

---

*M.H. Quenouille to Fisher: 11 March* 1957

Recently at a meeting in Imperial College, D.V. Lindley put forward an example which he alleged showed that Fiducial Distributions could not be used with Bayes' theorem. I was not present at this meeting, but I have managed to borrow the enclosed note circulated by Lindley. I understand that it is due to be published in *Biometrika*.[1]

The other enclosed note indicates how I think Lindley's arguments should be refuted. I should be grateful for any comments or suggestions that you may have on this. Would you please return the Lindley note when you have finished with it as I have this on loan?



There is a second matter which I should like to take this opportunity of raising with you. In 1952, I started a book entitled 'Foundations of Statistical Thought' which I expect is to be published this year. At the time Dr. Yates was kind enough to look over and comment upon the chapters concerning fiducial inference. My attitude on this agrees with your own apart from one or two relatively minor points.

The recent publication of your own book [*SMSI*] has made necessary several alterations in the text as originally planned, and I am still carrying this out. However, I should be very grateful if you have time to look over and comment upon the manuscript as finally prepared for press. I should explain that the book[2] is intended to be a concise account, for students, of the methods and, in particular, ideas that are in current use in statistics.

P.S. On page 135 of your book, the Bessel function should be $I_0(aR) = J_0(aiR)$.

[1] Published eventually as Lindley, D.V. (1958). Fiducial distributions and Bayes's theorem. *J.R. Statist. Soc.* **B 20**, 102–7.

[2] Published as *The fundamentals of statistical reasoning*. Griffin, London (1958).

*Fisher to M.H. Quenouille: 12 March* 1957

I was very much interested to have your letter and its enclosures. There have been so many variations in the notation used for Bessel functions that I am not yet sure whether I have, as you say, introduced yet another variant, or whether I have, as I thought, been following one of the reputed or semi-accepted notations; not, I think, that it matters much so long as the function concerned is independently defined.

As regards what you say about fiducial inference and Bayes' theorem, I imagine that Lindley's paper was put forward after conferring with George Barnard, who in turn had been discussing some aspects of these matters with me. Originally Lindley was inclined to attack the whole concept of fiducial probability on a wide front. I put it to George that the particular example in Chapter V [*SMSI*], in which I had used a fiducial inference in a Bayes-like manner, was objectionable in the first instance because I had ignored the requirement emphasized in Chapter III of sufficient estimation, and also for a more subtle reason, which may not appeal to everybody, namely that Bayes' argument was based quite explicitly on an independent prior act of sampling, and not only on the statement of probability derivable from that act. In fact, however, the parent population, and the act of sampling from it, cannot be inferred from the probability statement, and therefore contain a logically independent ingredient, which is lacking when only the probability statements are available.

Otherwise, of course, after deriving a probability statement from the sample, we could infer something about the parent population, which would be irrational since our information concerns only one population, and beyond that, nothing about any parent population from which it might have been derived by sampling, a point which I tried to make on page 120[1] in the paragraph overlapping page 121, which is perhaps more intelligible if the misprint of 'or' for 'of', at the end of the fourth line, is corrected.

Anyway, I am replacing example 6 (page 123) by a more competent one proving the same point in the second edition, which has already been asked for, and I presume that by the time Lindley came to give the paper you refer to at Imperial College, he thought it wiser to confine his attention to an example which, presumably, he already knew I wish to retract.

Do you know if 'Neyman's accredited representative' in this country is still opposing Behrens' solution of the comparison of the means of two normal samples on the ground that on 'repeated samples from the same population' his test gives a significant result less frequently in some special cases of the composite hypothesis tested than in others, or than the level of significance would imply for a simple hypothesis?

You might like to see the replacement section I have drafted. Please send it back and do not let it be made a means of attacking my work, until it is published.

[1] *SMSI*, p. 124–5.

## *M.H. Quenouille to Fisher:* 28 *March* 1957

Thank you for your letter and comments. I am returning the draft of the replacement section which you kindly loaned me.

I am not sure about the reasons (in particular, the first) that you give for changing this section. It seems to me that if joint fiducial distributions are to be something other than the product of independent fiducial distributions, some step involving statistics that are not fully sufficient is required. My formula (1) seems to me to be useful (whether it is called Bayes' theorem or not) in that it serves to connect fiducial distributions based upon partial information with those based upon the whole. In particular, as might be expected where the difference between $x$ and $y$ lies in the number of independent observations $n$ and $m$ used in their calculation, the two integrals given in that formula are in the ratio $n : m$. Hence

$$f_d(\theta \mid z) = \frac{n+m}{m} \int_{z(x,y)=z} f_d(\theta \mid y) f_1(x \mid \theta)\, dx.$$

A particular instance where this might be used is in relating the fiducial distribution of $\sigma^2$, given $a = \bar{x} - \mu$, to that when $a$ is not given. Here

$$f_d(\sigma^2 | S^2) = \frac{n}{n-1} \int_{nS^2=(n-1)s^2+na^2} f_d(\sigma^2 \mid s^2) f_1(a^2 \mid \sigma^2)\, da^2.$$



This seems to me to provide a justification for saying that fiducial statements of $\sigma^2$, given $s^2$, are similar in character to statements of $\sigma^2$, given $S^2$.

If you could manage it, I should welcome an opportunity to discuss this and other matters with you. The enclosed notes written in 1953, (with which I don't fully agree now) still show a number of points of difficulty, the major one being whether a distinction should be drawn between fiducial limits of estimation and fiducial limits of significance. It seems to me that a good case exists for doing so.

## *Fisher to M.H. Quenouille:* 29 *March* 1957

I should be glad to discuss your recent note and your Private Report No. 2 on 'Problems involving restricted parameters' whenever you can make me a visit, as I hope you will sometime. From London it is not too inconvenient to come down perhaps and lunch with me, but for a longer stop I could, of course, put you up as my guest in Caius.

I believe all these problems are intrinsically simpler than has sometimes been thought, but that a few fundamental ideas have to be straightened out, if misapprehensions are not to intrude themselves.

My previous example of combining information at two levels was, I think, faulty in principle and in detail, and that any number of other simpler examples would illustrate better all that I have to say.[1] What seems to me now important, and what was indeed always obvious, is that a probability statement representing a certain type of partial knowledge, of the same kind as that which might have been derived from a previous sampling process, does not imply the historical reality of any such process, or of the bulk, or population, from which the sampling is supposed to have taken place. In fact, Bayes postulated more than probability *a priori*; he postulated a prior sampling process, and it was this that allowed him to combine the probabilities of two sampling operations by the product formula, for they were known to be independent. A probability statement, on the other hand, derived from observations is not independent of any subsequent observations, and the combination must be effected at the observational level, and not at the level of inferred probability statements.

Let me know when you are coming.

[1] See also Fisher's letter of 8 February 1958 to Barnard (p. 38).

---

## *Fisher to C.R. Rao:* 6 *February* 1956

I have just seen your *Sankhyā* paper[1] 'On Fisher's lower bound', and it will certainly do nothing but good to bring the discussion of consistency into the open. You may have seen, for example, Anscombe's review[2] of the 12th edition of *Statistical Methods* in which, while complaining that there is nothing new in this old book, he shows that he cannot recognize anything new to himself by confusing the definition of consistency, made explicit for the first time in the 12th edition, with a mere stipulation of lack of bias, which since it is not invariant for functional transformation of parameters has never had the least interest for me.

I was, however, surprised by the statement at the foot of page 333 that I 'was considering only analytic functions of frequencies'. I do not believe I introduced the word 'analytic', for it seems to me that the properties of *analytic* functions, as developed in the theory of the complex variable, are really quite irrelevant. I did stress, not as an interpretation of my earlier definitions, but as a result of later reflection, that we are concerned in estimating only ultimately with functions linear in the frequencies, though I do not see why the estimate should not be a non-analytic function of such a function.

The introduction of the word 'estimator', I suppose by Maurice Kendall, seems well fitted in this connection to confuse the discussion by slurring over the distinction between a method of estimation on one side, and on the other the particular estimates, or class of estimates, to which this method leads in a particular case or class of cases.

However I am only concerned now to make it clear that the properties of analytic functions are to my mind not important in the stipulation that an estimate should be a consistent one. The suggestion to the contrary does suggest, as indeed you do on page 334, that I have changed my mind in this matter, a suggestion that will give great pleasure in Berkeley. I must rely on you to check their exuberance about this.

[1] Kallianpur, G. and Rao, C.R. (1955). On Fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhyā* **15**, 331–42.
[2] *J.R. Statist. Soc.* A **118**, 486–7 (1955).

### *C.R. Rao to Fisher:* 10 *February* 1956

Thanks for your letter of 6 February 1956. Our main aim in writing the paper is to draw the attention of the Statisticians to your definition of consistency. We are sorry for the paras. at the foot of page 333 and the top of 334 of our paper. We are aware that you have not introduced the word analytic in your writings but I remember to have heard your mentioning about the analyticity of the statistic as a function of frequencies. Maybe it is based on a misunderstanding. We will publish a correction in the next issue of *Sankhyā*.

Of course, so far as the definition of consistency is concerned analyticity of the statistic is not needed. But it seemed necessary to assume continuity to prove probability convergence. We had to state this because there exist artificial examples of a Fisher consistent statistic which does not tend to the true value of the parameter in probability. We needed differentiability to establish lower limit to variance. Otherwise counter examples can be produced. We wanted to stress these two analytical properties.



### *Fisher to C.R. Rao:* 16 *February* 1956

Thank you for your letter. I am still not quite clear what your ideas are, e.g. why you feel it necessary to establish a limiting property for large samples, when that property is well defined for all samples of finite size. Is this not a very long way round? I do not think any counter-examples are even conceivable if the matter is treated solely from the point of view of finite samples, nor would it matter if they were, for misbehaviour at the limit would be a very academic peril to a worker whose examples are, and always will be, finite. However, I may have misunderstood you. Perhaps you will let me see, before it appears, the correction you propose in the next issue of *Sankhyā*.

### *Fisher to C.R. Rao:* 21 *February* 1956

I have read your letter of February 15th, with the note of correction to be published shortly in *Sankhyā*.[1] I believe its publication will minimize confusion, but that some writers will try to take advantage of the original passage.

[1] See *Sankhyā* **16**, 206 (1956).

### *C.R. Rao to Fisher:* 6 *March* 1956

Thank you for your letters of February 16 and 21, 1956. I am glad to have the opportunity of corresponding with you and getting my ideas clear about this problem. I realise that limiting property is not important. But I am worried if somebody defines a function in the sample space such that on the curve of expectation it has the value $\theta$, the value of the parameter, and arbitrary elsewhere. Strong consistency or what we have called Fisher consistency is satisfied but such a function defined above cannot be used as an estimate. The condition of continuity in the neighbourhood of the curve of expectation may be a reasonable restriction on the function. This implies the limiting property of convergence in probability which is used merely as an alternative condition and is somewhat weaker than strict continuity.

### *Fisher to C.R. Rao:* 20 *March* 1956

I am glad to have your letter of March 6th, though perhaps I do not fully grasp what you are saying. If you consider a generalized space representing class frequencies observed, possible observations will fall on the lattice points throughout this space.

A continuous curve will represent multiple expectations for different values of the parameter, and therefore different points on this expectation curve are identified with the values of the particular parametric function used (of course more generally we may have a $p$-fold expectation space corresponding with simultaneous estimates of $p$ parameters).

If a 'statistic' means a determined function of the observed values, this statistic may be the same, or very nearly the same, at different lattice points, and if it is expressed explicitly in terms of the observed frequencies, it will determine a space of appropriate dimensionality passing through and among the lattice points and cutting the curve of expectations at a point already identified with a particular value of $\theta$. This same value of $T$ is then assigned to the estimate common to the whole equi-statistical region, if the estimate is to be consistent. In fact the expectation curve is used to assign metrical values to any system of non-intersecting equi-statistical regions.

Consequently I am not sure what you mean when you say 'if somebody defines a function in the sample space such that on the curve of expectation it has the value $\theta$ and arbitrary elsewhere'.

Formally the problem is simplified by considering only functions linear in the observed frequencies represented by regions planar in the multi-dimensional coordinates, and intersecting only in the impossible regions in which some frequencies are negative. It is intuitively obvious that only one such system can be both '*consistent*', that is have its regions correctly labelled, and *efficient* in the sense of striking the expectation line orthogonally, when orthogonality is defined in a Poissonian sense in relation to the expected frequencies at the point of intersection.

The justification, which I think should be not a prior but a posterior consideration, for using demonstrations appropriate only to functions linear in the frequencies, is that all other functions lead to discontinuities in the statistics calculated, when minimal changes in continuous measurements lead two distinct values to coincide, for any accurate measurement must be regarded as a frequency of unity for the number of occurrences in an infinitesimal range of its continuum.

George Barnard has also been reading your paper with great interest.

### *C.R. Rao to Fisher:* 29 *December* 1961

I could find some time this morning to give a first attempt.[1] The results are as follows.

$u_1, u_2, u_3$: independent normal variables, zero mean and unit variance.

$v_1, v_2, v_3$: indep. $\chi^2$ variables, apart from constant multipliers, with $n-1$, $n-2$ and $n-3$ d.f.

Then there exist functions $f_i(x,y,z)$, $i = 1,2,3$, such that

$$u_1 = v_2 f_1(r) - v_1 f_1(\rho)$$
$$u_2 = v_3 f_2(r) - v_2 f_2(\rho)$$
$$u_3 = v_3 f_3(r) - v_1 f_3(\rho) + f_1(r) u_2^{\dagger}$$

† this term arises due to orthogonalisation of $u_2, u_3$.

$$f_1(r) = \frac{r_{12}}{\sqrt{1-r_{12}^2}} \text{ as in Fraser}^2$$



$$f_2(r) = \frac{r_{32.1}}{\sqrt{1-r_{32.1}^2}}$$

$$f_3(r) = \frac{r_{31.2}}{\sqrt{1-r_{31.2}^2}} \cdot \frac{1}{\sqrt{1-r_{12}^2}} = \frac{r_{31.2}}{\sqrt{1-R_{1.23}^2}}$$

$f(\rho)$ are corresponding functions of $\rho$.

This is capable of generalization but I don't know whether any simplification is achieved.

I have just started checking whether the fiducial distribution of $\rho$ obtained from Fraser's equation is the same as yours obtained by inversion from the distribution of $r$. The answer is yes by 'general reasoning' but I must work out the details.

I hope to give some time to the other problem of fiducial distribution of a future vector observation $(x_1, \ldots, x_p)$ given a sample of $n$ observations from a multivariate normal distribution with an unknown mean and covariance matrix. I mentioned in a paper about 8 years ago that this distribution is useful in classification problems but thought it would resemble Hotelling's distribution. Recent work of Cornish shows that there may be slight differences.

[1] See Fisher's correspondence at this time with Fraser (p. 106) and Sprott (p. 215).

[2] Fisher has written in the margin alongside these three equations, cot C, cot A and cot B/cot C, respectively.

### *Fisher to C.R. Rao:* 23 *March* 1962

Did Fraser decide to visit you in India after all? I do not know whether you have been thinking of the trivariate problem. If you have made any progress you will not need this, but if not it may start a train of thought.

We want three pivotals each varying monotonically in one parametric function. I suggest

(a) a function of $r_{23}$, $\rho_{23}$ such as Fraser's,

(b) a function of $R^2$ and $\pi_1^2$,

$$R^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}^2}, \quad \pi_1^2 = \frac{\rho_{12}^2 + \rho_{13}^2 - 2\rho_{12}\rho_{13}\rho_{23}}{1-\rho_{23}^2}.$$

It is known that the distribution of $R^2$ depends only on $\pi_1^2$. For $N$ odd the probability integral has been given explicitly in $(N-3)(N-1)/8$ terms. In my 1928 paper on the Multiple Correlation [*CP* 61] I also bring in $\psi$. I think we might define $\psi$ as

$$\tan^{-1}\frac{(r_{13}-r_{12})\sqrt{1+r_{23}}}{(r_{13}+r_{12})\sqrt{1-r_{23}}} - \tan^{-1}\frac{(\rho_{13}-\rho_{12})\sqrt{1+\rho_{23}}}{(\rho_{13}+\rho_{12})\sqrt{1-\rho_{23}}}$$

with the simultaneous distribution

$$\frac{\{(N-2)/2\}!\,(1-\rho_{23}^2)^{(N-1)/2}(1-r_{23}^2)^{(N-4)/2}}{\pi^{5/2}(N-3)!\,\{(N-5)/2\}!}\,dr_{23}\left(\frac{\partial}{\sin\theta_1\partial\theta_1}\right)^{N-2}\frac{\theta_1}{\sin\theta_1}$$

$$\times\ (1-\pi_1^2)^{(N-1)/2}(1-R^2)^{(N-5)/2}d(R^2)d\psi\int_{-\infty}^{\infty}\frac{dz}{(\cosh z-\pi_1 R\cos\psi)^{N-1}}$$

[where] $\cos\theta_1 = -r_{23}\rho_{23}$.
The first line is $dP$, marginal, when $P(r_{23}, \rho_{23})$ is the probability of the variate $r_{23}$ being less than any given value $r_{23}$.

This convinces me that a trivariate fiducial distribution *exists*. I doubt if I shall ever clean it up. You might find the introduction to the 1928 paper worth re-reading. It has been called obscure.

The human mind tends to think everything new to it, or from which it has anything to learn, obscure.

If it works the process would seem to generalize easily for stepping up from $t-1$ dimensions to $t$.

---

## *V.I. Romanovsky to Fisher:* 17 *March* 1931

I thank you very much for the copies of your and your pupils' papers.

Your paper on inverse probability is very interesting but I cannot wholly consent with it. Let us take the simplest case: an event of the unknown probability $p$ is observed $np'(0 < p' < 1)$ times in very large number $n$ of trials. Let $f(p)dp$ be *any* continued differential *a priori* distribution of $p$, then it can be shown (e.g. in my note in the *Comptes Rendus* of Paris on the *a posteriori* probabilities, for 1929, October) that

$$\int_{p'-\alpha M}^{p'+\alpha M} f(x)\,x^{np'}(1-x)^{n(1-p')}dx \Big/ \int_0^1 f(x)\,x^{np}(1-x)^{n(1-p')}dx$$

$(M = \sqrt{2p'(1-p')/n}$, $\alpha$ — arbitrary positive number) is asymptotically the same with

$$\int_{p'-\alpha M}^{p'+\alpha M} x^{np'}(1-x)^{n(1-p')}dx \Big/ \int_0^1 x^{np'}(1-x)^{n(1-p')}dx$$

and the latter expression corresponds to the supposition that $f(x) =$ Const. I am sure that in many more complex cases we shall have the analogous results and then the theory of inverse probability is saved for these cases provided we have sufficiently large numbers of corresponding experiments. But it must be confessed that in the case of small samples this theory fails: then we cannot get rid of the functions like $f(p)$.



Moreover I think that your maximum likelihood is, if not logically, analytically very closely connected with the inverse probability. Indeed, the equivalence of the above written expressions and of all other analogons [*sic*] of them is based on the existence of certain maximums.

In conclusion I shall communicate you a theorem recently demonstrated by me.

Let $x_1^{(1)}, x_2^{(1)}, \ldots$ be a series of independent random variables such that $E\{x_i^{(1)}\} = 0$, $E\{x_i^{(1)2}\} = \sigma_x^2 = \text{const.}$, $E\{x_i^{(1)}x_{i+1}^{(1)}\} = 0$ $(i = 1, 2, \ldots)$, $E$ being the sign of mathematical expectation. Consider

$$x_i^{(2)} = x_i^{(1)} + x_{i+1}^{(1)} + \ldots + x_{i+s-1}^{(1)}$$
$$x_i^{(3)} = x_i^{(2)} + x_{i+1}^{(2)} + \ldots + x_{i+s-1}^{(2)}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$y_i \equiv x_i^{(n)} = x_i^{(n-1)} + x_{i+1}^{(n-1)} + \ldots x_{i+s-1}^{(n-1)}$$

($s$ — any integer $\geqq 2$) and

$$z_i = \Delta^m y_i = z_{i+m} - C_m^1 z_{i+m-1} + \ldots + (-1)^m z_i.$$

Then, if $m/n \to \alpha \neq 1$ as $n \to \infty$, the series $z_1, z_2, \ldots$ is obeying to a limiting sinus law with the period $L$ defined by the equation $\cos(2\pi/L) = r_1$, $r_1$ being the correlation coefficient of $z_i$ and $z_{i+1}$, that is to say, for $n$ sufficiently great, with the probability arbitrarily near to 1, the series $z_1, z_2, \ldots$ will follow the above mentioned sinusoide as near and during as many consecutive periods as you wish.

This theorem is the generalisation of the special case ($s = 2$) demonstrated by Mr. E. Slutsky (*C. R.* of Paris, t.185, p. 169) and costed me much efforts.

It is evident that, with aid of this theorem, we can construct such series of dependent random variables that will follow an arbitrary prescribed limiting harmonic law. I find this most remarkable!

It is worth to mention that the stochastical scheme lying in the base of the series $z_1, z_2, \ldots$ is a simple discrete or continuous chain of Markoff. . . .

## *Fisher to V.I. Romanovsky:* 31 *March* 1931

Thanks for your interesting letter of March 17th. I fully admit the truth in the limit $n \to \infty$ of the expression you give, but I do feel the difficulty that unless we set an upper limit to $d(\log f)/dp$ in the neighbourhood of the population value of $p$, we cannot say that our sample is large enough for the inverse inference to be justified. Sometimes certainly no such limit can be postulated. For example, the intraclass correlation for fraternities of $k$ cannot be less than $-1/(k-1)$; hence unless there is a law of nature that $k$ cannot exceed some determinate limit, the intraclass correlation in the population cannot be negative at all. But it may very well be zero. There is thus in this case a sharp discontinuity in $f$, and I feel doubtful whether in other cases, though we cannot prove the existence of the discontinuity, we have any right to assume its non-existence.

When the Mendelian ratios were discovered Mendel had many families giving approximately the ratio 3:1; the observations were equally compatible with a theory that the ratio was a transcendental number sufficiently close to 3. His inference that the true ratio was just the whole number must have been based upon the view that *a priori* the whole number ratios were more probable than the irrational ratios in their neighbourhood. Here again $f$ is discontinuous.

Your new theorem strikes me as very remarkable, and I hope you will be publishing the whole investigation; at present, I do not fully understand it, but will be sure to return to the matter, as soon as I have less correspondence to deal with. . . .

Click here for next section