way that will compel the attention of some of those who have previously disregarded it. Am I right, though, in thinking that their treatment of fiducial inference is at present limited to certain classes of distributions and parameters and does not embrace all the uses that you have made? I got the impression from their first paper that they had not subsequently read more, and I am not sure how generally they expect to be able to produce suitable pivotal functions. . . .

Fisher to D.J. Finney: 3 *May* 1962

. . . I do not suppose that Rao, Fraser and Sprott have cracked the inversion problem, but I suppose the direct simultaneous distribution will appear in *Sankhyā* shortly. I hope also the RSS may soon get around, after more than a year, to publishing my paper on Bayes' experimental procedure [*CP* 289].

I think Fraser must have started with some rather over-simple notions in which the phrase 'continuous group of transformations' was prominent, and exhaustiveness and monotonicity not very visible. Probably he had given little attention to Estimation. Still he extricated himself from the Tukey–Savage sort of rubbish. I do not know whether he has made any *amenda* about the correlation coefficient, in print at least.

My chief intention in the last example in *Statistical Inference* was to illustrate the existence of 3 strata of parameters and their corresponding statistical estimates. This rather complicates the exact statement of the conditions for simultaneous fiducial distributions. I believe I said something about it in my last letter, but I forget how much.

Fisher to D.A.S Fraser: mid November 1961

It was good to see your paper in the *Annals*[1] for that Journal needs the injection of a little sense and relevance. But I was sorry that you had let Tukey and Savage waste your time for those two able minds are themselves in such a mass of confusion and contradiction that they can scarcely fail to confuse and frustrate others. The last section of your paper seems to lack confidence. What to me needs clarification are such phrases as 'the frequency interpretation that customarily goes with confidence intervals'.

Do you mean, for example: — This interval calculable from the data will cover the true value in $(1-\alpha)$ of repeated random trials?

The probability that the true value lies in this interval is $(1-\alpha)$?

I gather the latter is unorthodox among the great herd of teachers in American mathematical departments, and it is certainly not a valid inference from a test of significance only. It is orthodox also to avoid questions of Sufficiency, although their relevance was pointed out quite 25 years ago.

Still how can the former version above be called an 'interpretation'? It altogether avoids the specification of uncertainty, and makes no specific inference. It seems to ride on the shadow of the fiducial inference, which is often rigorously valid for the same interval (in very simple cases).

I am glad you see the sense of Behrens' test. I am sending a couple of offprints from *Sankhyā* but they will take time travelling about the world.

[1] Fraser, D.A.S. (1961). On fiducial inference. *Ann. Math. Stat.* **32**, 661–76.

D.A.S. Fraser to Fisher: 24 *November* 1961

Thank you very much for your letter and the kind comments it contained. . . .

The section on the correlation coefficient — you made reference to it — contains a regretful error. The fiducial distribution for ρ obtained by a group-theory analysis involving regression of y on x, as obtained similarly based on regression of x on y, and the original fiducial distribution as defined in your paper 'Inverse probability' [*CP* 84] are it seems identical and correspond to the equation

$$w = v\frac{r}{\sqrt{1-r^2}} - u\frac{\rho}{\sqrt{1-\rho^2}}$$

where w, u, v are independent and are distributed as a standard normal variable, a χ variable on $n-1$ d.f., and a χ variable on $n-2$ d.f. Solving for r yields the ordinary distribution of r given ρ; and solving for ρ yields the fiducial (as determined by any of the 3 methods) for ρ given r.

Also, for u, v given, the relationship of r to ρ is essentially of transformation parameter form. Thus the marginal for r with respect to ρ or ρ with respect to r is an average of transformation relationships . . .

Fisher to D.A.S Fraser: 29 *November* 1961

Thanks for your letter and for letting me know that you have now removed the cause of confusion which appeared in your discussion of the correlation coefficient.

The form you send me

$$w = v\frac{r}{\sqrt{1-r^2}} - u\frac{\rho}{\sqrt{1-\rho^2}}$$

is most interesting and suggestive. I wonder if it is capable of generalization to the trivariate case; for this I have found it possible to express the distribution of r_1, r_2, r_3 in terms of ρ_1, ρ_2, ρ_3 only. As you would expect the distribution involves the determinant of the r to one power $(N-5)/2$ and that of ρ to another $(N-1)/2$ together with a well defined function of r and ρ, which for 3 variates can be written

$$\int_0^\infty\int_0^\infty\int_0^\infty (uvw)^{(N-2)/2}\exp -\tfrac{1}{2}(u^2+v^2+w^2-2\gamma_1 vw-2\gamma_2 uw-2\gamma_3 uv)\,du\,dv\,dw,$$

say $F_{N-2}(\gamma_1,\gamma_2,\gamma_3)$, where $\gamma_1 = r(\rho_1-\rho_2\rho_3)/\sqrt{(1-\rho_2^2)(1-\rho_3^2)}$, writing from memory.[1] The distribution is easily generalized for t dimensions. I do not know why Sam Wilks could not do it.

You can imagine that an inversion would be of great interest.

It was not until about 1957 that I realized how completely dumb University teaching in Math Stat was in the U.S.. I could not believe it at first!

[1] cf. *CP* 288.

Fisher to D.A.S. Fraser: 18 *December* 1961

As I have not heard from you, it may be that my reply to your letter of Nov 24 has gone astray. I think I did include the Calcutta address as above.

With respect to the very elegant formula of your letter, on trying to reproduce it, I am led to put to you the possibility that it should be

$$w = v\sqrt{\frac{n-1}{n-2}}\,\frac{r}{\sqrt{1-r^2}} - u\frac{\rho}{\sqrt{1-\rho^2}}$$

where, as in your formula,

w is normally distributed about zero with unit variance,
v is a random variable χ for $n-2$ d.f.,
u is a random variable χ for $n-1$ d.f.,

and u, v, w all independent.

I put the question whether this same approach would not give the simultaneous distribution of $\rho_{12}, \rho_{23}, \rho_{31}$ in terms of r_{12}, r_{23}, r_{31} observed.

I believe it should do so, but the approach is yours, and I think you should consider its extension.

D.A.S. Fraser to Fisher: 6 *January* 1962

Thank you for your letters of Nov 29 and Dec 18. . . .

Concerning the pivotal equation for the correlation coef.

$$z = \chi_{n-2}\frac{r}{\sqrt{1-r^2}} - \chi_{n-1}\frac{\rho}{\sqrt{1-\rho^2}},$$

this was derived by the group theory approach and made use of the canonical form for the Wishart distribution which is concealed in Mauldon *JRSS*, *B*17

(1955). For a sample of $n-1$ from a bivariate normal with means equal to zero it yields:

$$\begin{pmatrix}\sqrt{\Sigma x^2} & \Sigma xy/\sqrt{\Sigma x^2}\\ 0 & \sqrt{\Sigma(y-bx)^2}\end{pmatrix} = \begin{pmatrix}\chi_{n-1} & z\\ 0 & \chi_{n-2}\end{pmatrix}\begin{pmatrix}\sigma_1 & \sigma_2\rho\\ 0 & \sigma_2\sqrt{1-\rho^2}\end{pmatrix}.$$

This is based on positive upper triangular factoring of the positive definite matrices. The diagonal terms produce the equation:

$$\begin{pmatrix}\sqrt{\Sigma x^2} & 0\\ 0 & \sqrt{\Sigma(y-bx)^2}\end{pmatrix} = \begin{pmatrix}\chi_{n-1}\sigma_1 & 0\\ 0 & \chi_{n-2}\sigma_2\sqrt{1-\rho^2}\end{pmatrix}.$$

Taking the inverses of these diagonal matrices and multiplying on the right in the first equation produces:

$$\begin{pmatrix}1 & \Sigma xy/\{\sqrt{\Sigma x^2}\sqrt{\Sigma(y-bx)^2}\}\\ 0 & 1\end{pmatrix} = \begin{pmatrix}\chi_{n-1} & z\\ 0 & \chi_{n-2}\end{pmatrix}\begin{pmatrix}\chi_{n-1}^{-1} & \chi_{n-2}^{-1}\rho/\sqrt{1-\rho^2}\\ 0 & \chi_{n-2}^{-1}\end{pmatrix}$$

which yields

$$\frac{r}{\sqrt{1-r^2}} = \chi_{n-1}\chi_{n-2}^{-1}\frac{\rho}{\sqrt{1-\rho^2}} + z\chi_{n-2}^{-1}.$$

This rearranged is the pivotal equation.

A student and colleague of mine at Toronto is working on aspects of the higher dimensional case. The distribution of the r's can be expressed in terms of the ρ's in a pivotal form. This form can be inverted but certain symmetries may be lost because of the use of the partial correlations which are natural in the regression-group form. I write from memory and a little geometrical intuition.

$$\begin{pmatrix}S_1 & S_2 r_{12} & S_3 r_{13}\\ 0 & S_2\sqrt{1-r_{12}^2} & S_3\sqrt{1-r_{13}^2}\cdot r_{23.1}\\ 0 & 0 & S_3\sqrt{1-r_{13}^2}\sqrt{1-r_{23.1}^2}\end{pmatrix}$$

$$= \begin{pmatrix}\chi_{n-1} & z_1 & z_2\\ 0 & \chi_{n-2} & z_3\\ 0 & 0 & \chi_{n-3}\end{pmatrix}\begin{pmatrix}\sigma_1 & \sigma_2\rho_{12} & \sigma_3\rho_{13}\\ 0 & \sigma_2\sqrt{1-\rho_{12}^2} & \sigma_3\sqrt{1-\rho_{13}^2}\cdot\rho_{23.1}\\ 0 & 0 & \sigma_3\sqrt{1-\rho_{13}^2}\sqrt{1-\rho_{23.1}^2}\end{pmatrix}$$

$S_1^2 = \Sigma x_1^2$, $S_2^2 = \Sigma x_2^2$. The diagonal terms can be used to form diagonal matrices. The inverse of these diagonal matrices multiplying into the above equation on the right will yield a matrix equation which contains implicitly the marginal fiducial (group method) distribution for the ρ's in terms of the pivotal variables. . . .

The general Wishart distribution is available from the equation above. The equation gives a parameter transform of the distribution in the standardized case. The standardized case has its distribution available by a simple geometrical-regression analysis argument: $(x_{11}, \ldots x_{1,n-1})$ has length with a χ_{n-1} distribution. $(x_{21}, \ldots x_{2,n-1})$ has a component in the direction $(x_{11}, \ldots x_{1,n-1})$ and this component has a standardized normal distribution; it has a component perpendicular to this with length having a χ_{n-2} distribution. And so on for trivariate and multivariate case — components in the orthogonal frame built up by earlier components have standardized normal distributions — the remainder has a χ distribution.

On November 28th I gave a fiducial talk at Berkeley and it was received with utter contempt by Neyman and bare tolerance by the others. . . .

D.A.S. Fraser to Fisher: 10 *January* 1962

I hope my letter of Jan 6 reached you in India. I have now heard from Keith Hastings at Toronto and shall quote several of his formulae — they tie in with those in my Jan 6 letter. [With] $g_{ii} = \chi_{n-i}$, $g_{ij} = N(0,1)$, [$i \neq j$, and all g_{ij}] independent,

$$\begin{pmatrix} s_1 & s_2r_{12} & s_3r_{13} \\ 0 & s_{2.1} & s_{3.1}r_{23.1} \\ 0 & 0 & s_{3.12} \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ 0 & g_{22} & g_{23} \\ 0 & 0 & g_{33} \end{pmatrix} \begin{pmatrix} \sigma_1 & \sigma_2\rho_{12} & \sigma_3\rho_{13} \\ 0 & \sigma_{2.1} & \sigma_{3.1}\rho_{23.1} \\ 0 & 0 & \sigma_{3.12} \end{pmatrix}$$

[i.e.] $\mathbf{S} = \mathbf{G}\Sigma$ say. Let $\mathbf{D}(\mathbf{S})$ for example be:

$$\begin{pmatrix} s_1 & 0 & 0 \\ 0 & s_{2.1} & 0 \\ 0 & 0 & s_{3.12} \end{pmatrix}$$

The pivotal equation is $\Sigma\mathbf{D}^{-1}(\Sigma) = \mathbf{G}^{-1}\mathbf{S}\mathbf{D}^{-1}(\mathbf{S})\,\mathbf{D}(\mathbf{G})$ which yields

$$\frac{s_2r_{12}}{s_{2.1}}g_{22} = \frac{\sigma_2\rho_{12}}{\sigma_{2.1}}g_{11} + g_{12},$$

$$\frac{s_3r_{13}}{s_{3.12}}g_{33} = \frac{\sigma_3\rho_{13}}{\sigma_{3.12}}g_{11} + \frac{\sigma_{3.1}\rho_{23.1}}{\sigma_{3.12}}g_{12} + g_{13},$$

$$\frac{s_{3.1}r_{23.1}}{s_{3.12}}g_{33} = \frac{\sigma_{3.1}\rho_{23.1}}{\sigma_{3.12}}g_{22} + g_{23}.$$

Hastings then quotes the following probability density function for r_{12}, r_{13}, r_{23}:

$$K\,D_r^{-(n+3)/2}\,D_\rho^{-(n-1)}\{(1-\rho_{12}^2)(1-\rho_{13}^2)(1-\rho_{23}^2)\}^{(n-1)/2} \times \int\int_0^\infty\int (uvw)^{n-2} \times$$

$$\exp\{-\tfrac{1}{2}(u^2+v^2+w^2-2uvr_{12}\rho_{12.3}-2vwr_{23}\rho_{23.1}-2uwr_{13}\rho_{13.2})\}\,du\,dv\,dw,$$

$$\text{where } D_r = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{vmatrix}$$

and this has the form in your Nov 29 letter with a difference in power for (uvw).
Solving the pivotal equation the other way for the ρ's yields acc. to his letter: p.d.f. for $\rho_{12}\,\rho_{13}\,\rho_{23}$

$$K\,D_r^{-(n-3)/2}\,(1-r_{12}^2)^{-1}\,\{(1-\rho_{12}^2)(1-\rho_{13}^2)(1-\rho_{23}^2)\}^{(n-2)/2} \times$$

$$D_\rho^{-(n+1)}\,(1-\rho_{12}^2)^{-1}(1-\rho_{13.2}^2)^{-1}\int\int_0^\infty\int u^n v^{n-2} w^{n-4}\exp\{\text{as above}\}\,du\,dv\,dw.$$

This seems asymmetric generally; definitely is for $r_{12} = r_{13} = r_{23}$. This distribution relates the ρ's to the r's by reference to a definite order in which regression is run on the variables 1,2,3. . . .

Fisher to D.A.S Fraser: 11 *January* 1962

I have just seen your letter of Jan 6. Do not forget to look up Walter Bodmer, who has also had some experience of being 'bawled down' by Neymanians. They intimidate Americans successfully enough, especially refugees anxious to get posts in American Universities. I do not think they need intimidate anyone else. For your encouragement I transcribe the first paragraph of a letter just received from a mathematical logician[1] at the Rockefeller Institute.

'I have just finished your book *Statistical Methods and Scientific Inference.* I wish I had seen it sooner. I am just delighted with it, for I have felt altogether alone in my dire suspicions of the logical inadequacy of the theory of statistical inference as expounded by Neyman, Pearson and nearly every other statistician I can think of . . .'. Later on, he says 'I am interested in tracking down everything that could possibly be of use to me concerning fiducial inference; for this seems to me to be the fundamental form of statistical inference — the decision theoretic approach being etc.'

I think I told you I had the simultaneous distribution of r_{ij} given ρ_{ij} for t variates. It can be written (if I have room)

$$\frac{\pi^{-t(t-1)/4}2^{-t(N-3)/2}}{\{(N-3)/2\}!\ldots\{(N-t-2)/2\}!}\,|\rho_{ij}^*|^{(N-1)/2}|r_{ij}|^{(N-t-2)/2}\,dr_{ij}F_{N-2}(\gamma_{ij})$$

ρ_{ij}^* is the cosine of the dihedral angle of the ρ figure, or of the side of the polar figure.

$$F_{N-2} = \int_0^\infty \dots \int_0^\infty (u_1 \dots u_t)^{N-2} \exp\{-\tfrac{1}{2}(u_i^2 + \dots + 2\gamma_{ij}u_iu_j + \dots)\}[du_1 \dots du_t]$$

a useful function of γ_{ij} only [where] $\gamma_{ij} = r_{ij}\rho_{ij}^*$. All very compact and handy if put this way [cf. *CP* 288].

Starting with the distribution of ρ_{12} marginal, I should like you to get that of ρ_{13}, ρ_{23} conditional (a), then ρ_{14}, ρ_{24}, ρ_{34} conditional (b) and so on.

But of course we are thinking of the problem in rather different terms.

Rao here, who is quite first class, derives from your compact formula one for ρ which I like

$$\frac{1}{\pi(N-3)!}(1-\rho^2)^{(N-3)/2}(1-r^2)^{(N-2)/2}d\rho \int_0^\infty\int_0^\infty (uv)^{N-3}(1+\gamma uv) \times \exp\{-\tfrac{1}{2}(u^2+2\gamma uv+v^2)\}du dv$$

[where] $\gamma = -r\rho$.

Thanks for all the trouble you have taken. Would you consider a joint paper with Rao capping what I have done, with everything you two can do — for *Sankhyā*?

[1] H.E. Kyburg. For Fisher's reply to Kyburg's letter, see p. 187.

Fisher to D.A.S Fraser: 14 *January* 1962

You still tell me nothing about Walter Bodmer; won't you ring up the medical school and locate him? I am sure you will like him.

I feel sure your last two letters are a step back from the vantage point you had reached with your equation between r and ρ with $s_1, s_2, \sigma_1, \sigma_2$ eliminated. The general linear transformations in three dimensions do not transform spheres into spheres, or spherical triangles into spherical triangles unless they are rotations, and simple rotations do nothing.

I want to end my book [*SMSI*] (when it comes to a third edition) by saying that the right way of treating equation (231) had since been demonstrated by you, eliminating s and σ, in the form

$$w = \frac{r}{\sqrt{1-r^2}}\chi_{N-2} - \frac{\rho}{\sqrt{1-\rho^2}}\chi_{N-1}$$

from which for any chosen ρ the distribution of r is found as I gave it in 1915 and for any given r the distribution of ρ is what I gave in 1930.

That from it you and Rao had derived new forms of which one, parallel to 216, might be set out. Is it

$$\frac{\pi}{(N-3)!}(1-r^2)^{(N-2)/2}(1-\rho^2)^{(N-3)/2}d\rho\left\{\frac{\partial}{\partial(r\rho)}\right\}^{N-3}\left(N-2+\frac{\partial}{\partial(r\rho)}\right)\frac{\theta}{\sin\theta}$$

or what? I am scribbling from memory. Tell me what you make of it and I shall know what to put. Perhaps the last operand is $(\theta - \tfrac{1}{2}\sin 2\theta)/\sin^3\theta$.

It is not axiomatic that a triple or multiple distribution for ρ_{ij} exists. I guess it does, and that the only satisfactory demonstration is to take the variates in arbitrary order

$$\begin{array}{ccc} & \rho_{12} & \\ \rho_{13} & & \rho_{23} \\ \rho_{14} & \rho_{24} & \rho_{34} \end{array}$$

and exhibit *symmetrical* formulae emerging. Will not the partial regression of z on x and y, conditional on a known distribution of ρ_{12}, give a symmetrical result?

Of course I should love to see the De Lurys again. I must consult in Adelaide before promising anything specific. Remember I am aged and pretty blind, but still *compos mentis*.

Do not forget Walter.

[P.S.] For your approval: proposed addition to *Stat. Meth. and Sci. Inf.*, p. 175.

The correct way of using the facts stated in (231) has been more recently demonstrated by Donald Fraser and David Sprott who eliminate s_1, s_2 and σ_1, σ_2 obtaining the equation in three random variables

$$w = \frac{r}{\sqrt{1-r^2}}\chi_{N-2} - \frac{\rho}{\sqrt{1-\rho^2}}\chi_{N-1} \qquad (233)$$

in which w is a normal variable with mean zero and unit variance, while the two others are χ-variables with $(N-2)$ and $(N-1)$ degrees of freedom respectively. Assigning any value to ρ the equation gives the distribution of r as it was given in my paper of 1915, while assigning any value to r, it gives the fiducial distribution of ρ as first given in 1930, and as used in this chapter.

Explicit forms for the distribution of ρ have been derived by C.R. Rao from Fraser's formula. For example the form corresponding with (216) is

$$\frac{1}{\pi(N-3)!}(1-\rho^2)^{(N-3)/2}(1-r^2)^{(N-2)/2}\frac{\partial^{N-3}}{\partial(\rho r)^{N-3}}\left\{\frac{\theta-\tfrac{1}{2}\sin 2\theta}{\sin^3\theta}\right\}d\rho \qquad (234)$$

D.A.S. Fraser to Fisher: 27 *January* 1962

In your January 11 letter you ask if I would consider a joint paper with you and Rao. I would be honoured and pleased to collaborate. . . .

The transformation approach to fiducial came to definite form with me two years ago and led to a talk in August 1960 here at Stanford to the Inst. of Math. Stat. That talk was written down to form the article in the *Annals*.[1] In preparing for that talk I only sketched the details of the application of transformation to the bivariate normal and in the hurry incorrectly inferred that the 3 marginal distributions for ρ were not the same. David Sprott at Waterloo Univ. took time to check the details and let me know in October last that the three seemed to be identical; he quoted a triple integral with a relationship between variables, which relationship becomes the equation

$$z = \frac{r}{\sqrt{1-r^2}}\,\chi_{n-2} - \frac{\rho}{\sqrt{1-\rho^2}}\,\chi_{n-1},$$

upon interpretation. My method gives the formula which I have interpreted, but in working with it I shall need to acknowledge Sprott's checking of the details on the derivation and picking up the incorrect statement in my *Annals* paper.

My transformation method using Mauldon's results gives a quick derivation of trivariate and k-variate correlation distributions and might be used in the proof of the trivariate and k-variate distributions you report in your letters of Nov 29 and Jan 11. The various approaches to fiducial distributions for correlation coefficients fascinate me and I feel that I can contribute to the analysis. Unfortunately I don't see time for this before the end of March. Is all this consonant with a joint paper with you and Prof. Rao?

You mention the changes for the third edition of your book [*SMSI*] and my transformation analysis which handles equation (231) and for which simple elimination in your equations yields the equation

$$z = \frac{r}{\sqrt{1-r^2}}\,\chi_{n-2} - \frac{\rho}{\sqrt{1-\rho^2}}\,\chi_{n-1}.$$

I would be pleased to receive acknowledgment for this. For noting that your marginal distribution for ρ is implicit in this, contrary to the statement in my *Annals* paper, Dave Sprott's name should be mentioned.

You record an interesting distribution for ρ that Rao has worked with — I have no preferences at present for one form over another — you have magnificent facility for handling such expressions.

The details of the successive analysis of the ρ's intrigues me — I at the moment don't see time to go into the details at this hectic and hurried place, Stanford, during this quarter. My colleague Hastings is working in this area and I shall certainly refer to you any results for this analysis. He claims in his last letter to have several other groups for the plane that yield symmetric fiducials. I think he is breaking with a certain aspect of my transformation approach — and getting interesting results — but at the moment I don't have clear feelings on their relationship to my work so far. . . .

[1] See p. 107.

Fisher to D.A.S. Fraser: 2 *February* 1962

I have your letter of Jan 27th. I am delighted that you think you see a straight path to the general inversion. May it not disappoint you.

The position about publication is this. *Sankhyā* has and is setting up my paper on the general r_{ij} distribution. In handing it in I said I did not much want to publish until you or someone else could follow it with the fiducial inversion. I had hoped that a paper by yourself and Rao might suit you both. He is an excellent collaborator, as well as first class in his own right.

You owe me an *amenda* for letting people think (as some are too eager to) that there was doubt as to the uniqueness of my solution of the bivariate case, which logically was elementary enough, and the most emphatic repudiation you can make of this misapprehension is to generalize it coherently with the multiple distribution I propose to publish.

It has also this importance: we know too little of the conditions in which fiducial distributions are to be expected. If exhaustive estimation is, as it may be, a sufficient condition, then $r_{ij} - \rho_{ij}$ is a test case. It is known not to be, strictly, a necessary condition, but it would be important to establish a group of sufficient conditions.

I expect to leave India before the middle of February. I believe you will be very welcome here if you can bear the hot weather expected by the end of March. No doubt Rao or Mahalanobis will write to make this clear.

Sprott has some difficulties about integrals based on my 5-parameter distribution, but I cannot think they are serious.

Send me a line on the strategy of the approach you think will work and give my love to Walter.

Fisher to D.A.S. Fraser: 13 *February* 1962

I thought I had made clear how pleased I was at the suggestion of my working for a year at Toronto. I expect De Lury will get around to making me a definite offer. I explained that I should have to discuss the matter in Adelaide. I should like to include the period of ISI meeting in Ottawa, which should be convenient.

I have heard from Sprott who also approves the wording I propose.

I shall be leaving for Adelaide on the 15th. I think you have my address there

Division of Math. Stat., CSIRO,
Adelaide University,
S. AUSTRALIA

I believe distributions, or equivalent general probability statements about functions of σ_1, σ_2, ρ can only be valid if based primarily on the marginal or unconditional distribution of ρ. Probably this has given Sprott and perhaps you some trouble. The individual statistics of an exhaustive set are not necessarily exhaustive for the corresponding parametric function.

However, you will have thought about all this.

Fisher to D.A.S. Fraser: 3 *May* 1962

I suppose you decided not to go to India after all. Have you had any useful correspondence with Rao lately or Sprott?

I heard from Dan and suggested coming about the end of March 1963 working (perhaps with advanced students chiefly) until the end of May, visiting Britain in June and July, and returning to Saskatchewan and later Ottawa in Aug and Sept; then serving a term until Christmas and perhaps spending that cheerful season with my daughter in Wisconsin. All this may not suit perfectly, but I hope it can be made to do.

You may remember my suggesting that the general inversion of the simultaneous correlation problem might be approached by steps. From the marginal bivariate, about which I hope and think there should be no further difficulties, one might graft on the formulation as in my 1928 paper on multiple correlation, the simultaneous distribution of ρ_{13}, ρ_{23} in an array with given r_{12}, ρ_{12}. You might well find the introductory section of the 1928 paper helpful, using ψ as pivotal instead of integrating it out as I did to find the distribution of R. I think if ρ is the true multiple correlation, $d\rho d\psi$ is

$$d\rho_{13}\, d\rho_{23} \;/\sqrt{1-r_{23}^2}$$

or something like it.

Manipulation of the full explicit expression is rather tough and here it may be that your use of exact random variables may provide the key.

With good wishes to Sprott and Dan De Lury.

Fisher to D.A.S. Fraser: 7 *May* 1962

I have just seen yours of 1st May. I am indeed hoping to see a good deal of you after you are free from the fairly congested term's work Jan–March. I could come earlier than the end of March if that is desired, but I imagine that you and a few of your brighter men may be more free for concentrated thought after March.

Sprott also, I understand, is in the same Province, and I should be willing to take quite a bit of trouble to see something of him too this time (April, May).

. . .

Have you contacted Henry Kyburg of the Rockefeller Institute, New York? I have not yet seen his book, which is expensive even for an American book (Wesleyan Univ Press) and may be large. I suppose it will obliterate Keynes' book of 1928, which also was ambitious, but in retrospect not very good.

Have you heard at all further from Rao? Even if not this year you must visit the Stat. Inst. in Calcutta some time, but I should advise Dec and Jan. Rao is always worth talking to.

It is serious that so many young men, like you and George Box, have had to learn to disregard the indoctrination pressed upon them at their Universities. A live centre of teaching so near as Toronto would certainly make a number of U.S. centres sit up and take notice. I think they induce the belief that they already know everything by attending too many seminars.

Fisher to D.A.S. Fraser: 22 *May* 1962

I delayed answering your letter of May the 9th to give Alf Cornish a chance to consider it. Now he has returned in some triumph with a 3 million pound computer, as it were, in his pocket, I can say that it looks O.K., and I am prepared to arrange to fly to Toronto at the end of April. Dan had better arrange through the airline, presumably Qantas, for a credit at that time.

About the correlation problem, it does involve I fancy further intricate analysis, of a kind which the classical analysts have largely ignored. I mean something like this; the distribution of r_{23} depends only on ρ_{23}. Therefore there is a pivotal involving only these two, the distribution of which is independent of ρ_{12}, ρ_{13}, ρ_{23}. Equally if R is the multiple correlation of (1) on (2,3) the distribution of R has been expressed as a function of ρ only, where ρ is the corresponding parameter. I would swear, though I would not undertake to demonstrate, that these two distributions are statistically independent, so there is a pair of pivotals. I doubt if it is mathematically possible that there is not a third, involving say

$$\frac{\rho_{12}-\rho_{13}}{\rho_{12}+\rho_{13}}\,,\quad \frac{r_{12}-r_{13}}{r_{12}+r_{13}}$$

to complete the specification of statistics in terms of r_{12}, r_{13}, r_{23} and of the parameters in terms of ρ_{12}, ρ_{13}, ρ_{23}. It could be that this third distribution has a twist and fails to be monotonic. I will believe that when I see it. However, those pivotals with a joint distribution independent of the parameters, and each monotonic in the appropriate parametric function must give the fiducial inversion. Probably one needs something less evil than the probability integrals of the first two distributions, which is why I was excited by your use of random functions in

$$\frac{r}{\sqrt{1-r^2}}\,\chi_{N-2}-\frac{\rho}{\sqrt{1-\rho^2}}\,\chi_{N-1}$$

and the real reason why I threw the problem at you. Let me know.

M. Fréchet to Fisher: 8 *January* 1940

May I consult you on two points (even three). It happens that in two ways I am much more engaged this year in mathematical statistics than I was ever before (when I was more active in other fields).

My colleague Darmois being mobilized, I replace him in teaching of math. statistics and besides I have been appointed Director of a so called 'Laboratoire' of Mathem. Statis. and Calculus of Probability organized in view of researches interesting National Defence.

As the Directory of this 'Laboratory', I will probably soon be sent to England to establish contact between statisticians working [on] both sides of the channel for war purposes. This letter is then first intended to beg you to let me know whether you are fully engaged in your usual so valuable researches or whether you have to divert some of your labour for special researches for National Defence. In the last case, I should like to have from now your views concerning the different fields where statistical researches could be done for war purposes and during my visit in England we might précise [*sic*] some details which it is better not to précise in letters. Even in the case where you are not in touch with military departments, you might perhaps let me know the names of mathematical statisticians engaged in work for these departments. So much for this part of my new work.

Concerning the other part, that is, my course on mathematical statistics, I had to look more attentively than I had time to do before, on some new developments of statistical theory. And I would like to have your opinion on the following quotation of a paper by Deming and Birge[1] 'On the statistical theory of errors'. This is page 142, end of first column and beginning of the second. 'These particular values of σ and s are accordingly so related to each other that if σ were actually the S.D. of the parent population then there would be 19 chances in 20 that a sample drawn therefrom would have a S.D. as large as or larger than s; and conversely, since s has actually been observed, there is only 1 chance in 20 that the S.D. of the parent population is as large as or larger than σ.'

They have added a slip to introduce some corrections to their paper, but these quoted lines do not seem to have been corrected. I have underlined 'and conversely' because it seems to me that a logical confusion has there arisen. If I understand well, the probability mentioned there is computed from equation (30):

$$P_s = 1 - \Gamma_\nu\left(\frac{n-1}{2}\right) \Big/ \Gamma\left(\frac{n-1}{2}\right); \; \nu = ns^2/2\sigma^2.$$

This equation holds good when σ *being fixed*, s is a *random* variable computed from the observed values of the random variables $x_1, \ldots x_n$. Now in the second part of the quotation, the former value of this probability (computed *before the trials*) is considered also as valid when, s being observed, σ is considered as a random variable. As the whole proof of the formula is based on the first hypothesis and not on the second one, I do not see why the formula should be still valid in these very different circumstances.

It is so much important to clear up the validity of these quotations that it seems that this logical confusion is often met, though in sentences which are not clear enough, in many modern papers on statistics.

P.S. Vous avez obtenu la loi de probabilité de $(\bar{x}-\mu)/s$ (avec vos notations). A-t-on obtenu et publié (et dans ce cas, où?) la loi de probabilité de $u = \{\gamma_1(x_1-\mu) + \ldots + \gamma_n(x_n-\mu)\}/s$ où les γ sont des constantes ($\Sigma\gamma_i = 0$), les x_i valeurs observées d'une variable gaussienne X, $s^2 = \Sigma(x_i-\mu)^2/(n-1)$, $\mu = \bar{X}$?

[1] Deming, W.E. and Birge, R.T. (1934). *Rev. Mod. Phys.* **6**, 119–61. See Fisher's letter of 25 September 1934 to Deming (p. 80).

Fisher to M. Fréchet: 17 *January* 1940

I am very glad to have your letter of January 8th with the quotation from Birge and Deming. I find these writers often very obscure, but as they are obviously influenced by the form of argument for which I am responsible, I should like to make my own position, at least, clear to you.

I enclose the paper on 'Inverse probability' [*CP* 84] in which I first introduced the fiducial argument, though, in fact, as my reference to M. Ezekiel shows, many people had been arguing in this way from the moment when theoretical distributions, such as χ^2, t and z, were first tabulated so as to show the values taken at different levels of significance (values of P) instead of showing the values of P for different values of χ^2, etc.

I should like you to keep the offprint so long as it is useful to you; but, as I have very few copies, perhaps you will send it back if, at any time, you find you no longer want it. A second paper (The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proc. Roy. Soc. Lond. A, 139:* 343–348) [*CP* 102] also expresses well what is still my point of view. I have no copies of this, but the *Proceedings of the Royal Society* may be accessible to you.

With respect to your other mathematical question:—

If $x_1, \ldots x_n$ are values independently and normally distributed with standard deviation equal to σ, and if $S(Y) = 0$, then $S(Yx)/\sigma$ is normally distributed about zero with standard deviation equal to $\sqrt{S(Y^2)}$ and this distribution will be absolutely independent of that of

$$s^2 = S(x-\bar{x})^2/(n-1)$$

or of s/σ. So that the ratio of these two quantities

$$S(Yx)/s$$

will be distributed exactly as is $t\sqrt{S(Y^2)}$, where t is the value I tabulate for 'Student's' distribution.[1]

I emphasise this type of extension of 'Student's' original argument in a paper in *Metron* (Applications of 'Student's' distribution) 1925, Vol. V, Part 3: 90–104 [*CP* 43]. It is really the basis of the tests used in the Analysis of Variance.

On your mission to England I think you ought to see Professor Lennard-Jones, at the Mathematical Laboratory, Cambridge, who will be able to put you in touch with war work involving mathematical statistics better than any one else. My laboratory has not yet been used for this purpose.

[1] See also Fisher's letter of 26 April 1940 (p. 134).

M. Fréchet to Fisher: 21 January 1940

I have read with great interest your paper 'Inverse probability'. I understand all the paper and agree with many of its statements, but I must confess that I find here precisely the *same difficulty* as the one I pointed out in Deming and Birge's paper. The difficulty arises in a very short sentence which looks as obvious as the other ones but of which the meaning has a capital importance (as the words 'and conversely' in Deming and Birge):

Page 533: 'we may express the relationship by saying that the true value of θ will be less than the fiducial 5 per cent value corresponding to the observed value of *T in exactly 5 trials in 100*'. Now the five per cent and the 5 trials in 100 refer to two probabilities of the same event, it is true, but the populations where these probabilities are computed are extremely different. In the first one θ is fixed whereas T is a random variable and the table has been computed on this assumption. In the second one T is fixed, θ is a random variable and the first table which is still used (or the corresponding formula) has *not* been computed under this second assumption.

Such is my difficulty and I have found the same difficulty in other statistical papers.

In fact there is a long time [*sic*] I had doubts on such a point but they had [arisen] in reading sentences which were not definite enough to lead to something more than suspicion. The 'and conversely' was the first sentence where I found clearly an identity admitted as obvious between probabilities of the same event in two very different populations.

As you kindly allow me, I will keep sometime your paper to think more of the matter. I feel that it may be possible that the identification which I do not admit as obvious may still be used with success frequently in practice, though perhaps not always.

I am very glad to have got the reference to your paper in *Metron*; it will save me the time of doing laboriously what has been already brightly done.

My thanks also for mentioning Prof. Jones' work which will simplify my quest in England—where I am supposed to land in London 28th Jan. morning.

Fisher to M. Fréchet: 26 January 1940

Many thanks for your letter on Inverse and Fiducial probability, which is so clearly expressed that there is no difficulty in specifying exactly the point where our thought diverges. You say, in your third paragraph: 'In the second case T is fixed, θ is a random variable, etc.' If this were so, then the probability statement under consideration would be precisely a statement of inverse probability and, as such, as we both agree, probably not true, and certainly not known to be true.

I tried to make clear the difference in logical content between statements of inverse probability and statements of fiducial probability in my last paragraph, in which I specify: 'Whereas, however, the fiducial values are expected to be different in every case, and our probability statements are relative to such variability, the inverse probability statement is absolute in form, etc.'. From this it is, I hope, quite clear, that in the fiducial statement we are not considering T to be fixed. If, on the contrary, we remember that T will vary from sample to sample, and with it the corresponding fiducial 5% value of θ, then it will be true that the true value of θ will be less than the fiducial 5% value in exactly five trials in 100.

Some people prefer to make the same statement in the form of a disjunction, namely, either something has occurred which is known to occur in only 5% of trials, or θ exceeds $\theta_{5\%}(T)$. The important point, however, is that statements of fiducial probability have a logical content different from the more familiar statements of inverse probability, and are not intended to be equivalent to such statements. They refer, as you clearly perceive, to a different conceptual population. I have done my best, though perhaps not very successfully, to make this clear from my first references to the subject.

When in England I hope you may find time to visit me here in Harpenden, where my home is, and where I should be happy to have an opportunity of introducing you to my family.

M. Fréchet to Fisher: 3 February 1940

Arrived today in Paris, I find here your letter. After answering it on one point, I would like to submit to you a complement to a theorem of yours, in a different Chapter.

I. On the question of fiducial probability I would not like to let you lose your time were it not that the question of principle, there, is important and also that I heard recently of similar objections. Some were presented to me by a very good young French mathematician and some have appeared in a recent paper by Gini reproducing a speech 'I pericoli della Statistica' (*Rivista di Politica Economica*, 1939) which—as far as I understand this speech in *Italian*—looks very important. We are then at least three to have come, as it appears—to the same doubts, quite independently.

Now, it may be understood that, as you say, T will vary from sample to sample and with it the corresponding fiducial 5% value of θ. Still I expect that everybody will understand that the probability 5 trials in 100 is computed in the population where T gets a fixed value. You can see by Deming's quotation that at least Deming and Birge understood that, and other instances could be quoted.

But, even if we leave that, it remains that identification is admitted between probabilities of the same event in 2 populations which are different and essentially different because *in one, θ is fixed* (and this is implicitly supposed in the computation of the corresponding first probability) whereas *in the second, θ is supposed to be a random variable*.

The second way to state the meaning of fiducial probability, which is found in the second page of your letter, is essentially different and I quite agree with it. In fact it coincides with a statement which by chance came recently to my attention by Jan Wiśniewski: A note on inverse probability, p. 417 of the *Journal of the Royal Statist. Soc.* 1937:—

'either p is within the limits' (defined by some inequalities) 'or an improbable event happened, "improbable" to mean one whose probability is less than ' (5% in our case).

May I add that since I met the Deming's converse proposition, I noted also a converse proposition which appeared to me as of the same kind, in your paper 'The fiducial arg . . . .' *Annals of Eugenics*, vol. VI, 1935, at pages 391–392 [*CP* 125]: 'Since . . ., we may state the probability that μ is less than any assigned value'.

In order to assign this value $\mu_1 = \bar{x} - st_1/\sqrt{n}$, it is not sufficient to give the value of t_1 corresponding to 5 trials over 100, but also to give $\bar{x}$ and s or at least the linear function $\bar{x} - st_1/\sqrt{n}$ of $\bar{x}$ and s so that here we have really a hypothesis similar to the fixed T, that is, this linear function which was random in the proof of the law of t, whereas μ was fixed, becomes here fixed whereas μ is random in the statement on the law of μ.

Furthermore, I cannot reconcile the quotation by Wiśniewski of a statement in your paper in *Phil. Trans.* 1922, p. 327 [*CP* 18] (according to which it is impossible to get anything about the confidence interval of p from the information supplied by the sample) and the statement of the other more recent paper according to which the observation of $\bar{x}$ and s might give you the law of prob. of μ.

II. In your important paper 'The logic of inductive inference', *Journ. Stat. Soc.* 1935 [*CP* 124], you find an upper bound i of $1/(nV)$ and you prove that this very bound is obtained when choosing θ by the max. likelihood. It seems to me that, to be complete, the second part should also prove that this method gives a statistic T which is distributed normally about θ.

Perhaps you have given that proof elsewhere and in such a case I would beg you to give me the corresponding reference.*

However, I have discovered that this complement is not necessary *if* the first part of your theorem is generalized so as to get rid of the hypothesis of the normality of T (either for any n, or even only for great values of n). Now I thought it might interest you to hear that I have obtained such a result. Of necessity, my proof is absolutely different from yours where the hypothesis of normality fundamentally intervenes in the demonstration.

As our time was very short and we had to limit ourselves to meeting those engaged in war work, I could not think of trying to see you, at least this time. But I have, all the same, very much appreciated the courtesy of your kind invitation.

[P.S.] *I just now notice that Dugué proved this in his thesis under convenient hypotheses.

M. Fréchet to Fisher: 5 February 1940

I believe that the method I communicated to you in my last letter enables me to obtain something similar to the second theorem of your paper on 'The logic of inductive inference'. But I find it difficult to guess what is the exact meaning of your statement

'that for certain estimates, notably that arrived at by choosing those values of the parameters which maximize the likelihood function, the limiting value of $1/(nV) = i$.'

For the method of max. likelihood gives a value θ_0 of θ *after n* trials have given a sample; and in the formula, V is the variance of a random variable, which is a function $H(X_1, \ldots X_n)$, H appearing as a function chosen *before* the trials. So that the word estimates appears to cover in the same statement two significations: the numerical estimate θ_0 after the trials; the function $H(X_1, \ldots X_n)$ defined before the trials, (though its numerical values are known only after the trials). The result of my method leads me to a statement where I have to distinguish sharply between these two meanings.

Allow me to come to a second question, a historical one, based on your answer (p. 77 of the same paper) to the discussion. Is it right to say that the *formula* $1/(nV) \leq i$ was found by Filon and Pearson, or Edgeworth, but that they applied it incorrectly, in some cases? Is it right to say that some authors found that the equality $1/(nV) = i$ holds (at the limit) when the estimate is obtained through different (right or wrong) methods, but that none of them proved the same thing when the method employed is that of maximum likelihood, before your paper was published?

Finally a last question. In the same printed reply of yours, I read

'I mean by mathematical probab. only that objective quality of the individual which corresponds to frequency in the population, of which the individual is spoken of as a typical member.'

I have given a definition of probability which is different but which appears to me as having the same purpose. I wonder whether you would consider your definition, though perhaps not equivalent to mine, still as being at least consistent with mine. My definition is given in my books but may be summarized as follows:

The probability p of an event E in a category C of trials (in a universe U) is a physical magnitude $P(E/C)$ of which the frequency r/n of E in n trials of C is an experimental measure. . . .[1]

[1] For Fisher's reply, see his letter dated 12 February 1940.

Fisher to M. Fréchet: 10 *February* 1940

I doubt if we shall be able to get to a clearer understanding of the problem of fiducial probability, unless you are willing to accept it as a fact which I demonstrated to you by quotation in my last letter, that for the population of cases relative to which a fiducial probability is defined, the value of any relevant statistic T is not regarded as fixed. This I have deliberately exerted myself to make clear since my first writings on the subject. You will understand, therefore, that I find it a little disappointing when, on the third page of your letter of February 3rd, I find you saying: 'this linear function which was random in the proof of the law of t, whereas μ was fixed, becomes here fixed whereas μ is random in the statement of the law of μ'.

I shall be glad to give you all possible support in dissuading mathematicians from thinking that they can obtain a true probability statement logically equivalent to one of the kind aimed at by Bayes' theorem, yet without using the approximate basis of this theorem. Believe me, I have never attempted anything so foolish. The inferences which can be drawn without the aid of Bayes' axiom seem to me of great importance, and quite precisely defined, but are certainly not statements of the distribution of a parameter θ over its possible values in a population defined by random samples selected to give a fixed estimate T.

. . .

P.S. You say you cannot reconcile the point of view I have been expressing since 1930 with a too sweeping statement I made in the *Phil. Trans.* in 1922. The explanation is that fiducial probability was discovered in the interim. My 1922 statement was simply intended as a rejection of probability statement[s] based on the Laplacian principle of insufficient reason, which were the only statements concerning the probability of parameters made, I believe, up to that date. I did not then realise that statements of a logically different kind were possible, and rigorously deducible from the data. It is sad, but true, that every advance in Natural Science, including mathematics, means that some previous work is erroneous, inadequate, or obsolete.

Fisher to M. Fréchet: 12 *February* 1940

In my paper of 1922, *Phil. Trans.*, which I think you have available, I comment on the use by Pearson and Filon in 1898 of the formula giving the minimal variance or standard deviation of an estimated parameter.

If I remember right, they arrive at the formula by a very obscure argument involving inverse probability; but probably what they have done is equivalent to proving that $\sqrt{n}$ times the error of an estimate, taken e.g. approximately at the mode of an inverse probability distribution, will in the limit when n is large tend to be normally distributed with variance $1/i$. One of the features which makes their treatment so obscure and indefinite is that they do not notice that different methods of estimation have, in the limit for large samples, different precisions. Consequently they apply the formula, without hesitation, to estimates found by fitting Pearsonian curves by moments, and obtained a number of erroneous formulae, which were, I believe, not corrected till early in the present century when Sheppard showed how the standard errors of moments could be calculated directly. The erroneous formulae seem thereupon to have been dropped by the Pearsonian school, but there seems to be no hint in *Biometrika* which would inform the reader that previous misleading formulae were being corrected.

Edgeworth wrote a series of papers about 1908 in the Statistical Society's *Journal*. His attitude seems to be rather over-cautious than over-confident. He refers his readers to the paper by Pearson and Filon in terms which leave little doubt that he regarded it as correct, although in numerous other passages in this series of papers I should have thought he must be taken as recognising that different methods of estimation possessed different precisions, that these precisions have an upper bound, and that an estimate having in the limit the highest possible precision can be found by the procedure which Edgeworth regarded as inverse probability, but which, as I was concerned to emphasise in the 1930 paper I sent you, can and indeed must be completely dissociated from inverse probability, and which in 1922 I called the method of maximum likelihood.

The confusion of associating this method with Bayes' theorem seems to have been due originally to Gauss, who certainly recognised its merits as a method of estimation, though I do not know whether he proved anything definite about it.

I do not know of any explicit statement of the properties, consistency, efficiency and sufficiency, which may characterise estimates prior to my 1922 paper. I had noted the functional peculiarity of sufficiency in an early paper, 1920, in the *Monthly Notices of the Astronomical Society* [*CP* 12].

Yes, I think your definition of probability is consistent with mine, though I should emphasise that the physical magnitude considered is a character of the universe (population sampled).

M. Fréchet to Fisher: 20 *February* 1940

When you pointed my attention on the lines of your first paper on fiducial statem. where you state that T is not fixed, I noted this fact. I have not seen a similar warning in the paper which I quoted from *Annals of Eugenics*, 'On the fiducial arg.'. But since you say that when you wrote that second paper, you intended to mean that in the fiducial probability, $\bar{x}$, s as well as μ were not fixed, I have to admit that you know better than I do what you intended to mean.

Thereupon my question remains: on which ground are identified (or rather equalized) the two concerned probabilities referring to the same event (one particular inequality) but two different populations:

	Paper on inverse probability	Paper on the fiducial argument
1st population	T random, θ fixed	$x_1, \ldots x_n$ random, μ fixed
2nd population	T and θ random	$x_1, \ldots x_n$ and μ random

I mean, are those two probabilities equalized:

because it is obvious to you that they are equal?
or because it is a logical deduction of classical principles of Cal. of Prob., and then how?,
or because of the admission of a new principle and then which one?
or etc?

I thank you for the historical information which you gave me.

P.S. Trying to extend to small samples the second theorem of your paper: the logic of inductive inference, I find that in order that the estimate $T = H(X_1, \ldots X_n)$ which gives $1/(nV) = i$, should come from a function $H(x_1, \ldots x_n)$ independent of θ, it is necessary that f should be of a special form

$$f(x,\theta) = \exp\{\mu_0'(h(x)-\theta) + \mu(\theta) + g(x)\}$$

[with] μ, g, h arbitrary as long as they are such that $\int_{-\infty}^{+\infty} f\,dx = 1$. And then $T = \Sigma h(X_i)/n$. Among these functions f is for instance:

$$f(x,\theta) = (1/\sigma\sqrt{2\pi})\exp\{-(x-\theta)^2/(2\sigma^2)\},$$

more generally, $f(x,\theta) = (1/\sigma\sqrt{2\pi})\,h'(x)\exp\{-(h(x)-\theta)^2/(2\sigma^2)\}$ and so on. (It had seemed necessary to admit that $H(x_1, \ldots x_n)$ be independent of θ, to get a rational theory of estimates).

Fisher to M. Fréchet: 26 *February* 1940

Thanks for your note of February 20th. I think logically the fiducial argument proceeds in three stages, setting aside for the moment my usual cautions about using the whole of the information.

1) A continuous distribution is found for T for samples of a given size drawn from a population having parameter θ. θ is then also a parameter of this distribution of T.

2) A relation is established between the true value and any percentile point T_p of the distribution of T. We shall suppose that this also establishes a univalent inverse relationship from which, given T_p, θ may be found. It is then true for all samples of the given size (or otherwise specified by ancillary statistics) that the inequality T exceeds θ will occur with given frequency when T and θ are mutually related as defined above.

3) In these circumstances *I think it proper* to refer to p as the fiducial probability that θ is less than T. This as it stands is a definition of the phrase *fiducial probability*. I believe it is, properly speaking, a probability, measuring as it does the relative frequency of one out of two or more well defined outcomes of a well defined procedure. I think it may be described properly as the admission of a new principle, if this phrase means, as I suppose it does, the thinking of a given situation in an unfamiliar way. Alternatively, I have no objection to regarding stages 1) and 2) as logical deductions, and 3) as an arbitrary definition. The definition is, however, a matter of choice and not a matter of chance.

The outstanding difference from inverse probability lies in the population of events of which the particular one, to which the probability refers, is regarded as a member. In the case of inverse probability this population is that of all samples of a given size, selected to have a given value for the estimate T, drawn by chance from a population which has itself been drawn by chance from a super-population having a given specification in respect of the distribution of the parameter θ.

The population of events referred to in fiducial probability consists of all samples of a given size drawn from any population defined by some value or other, θ. It is obvious that the frequency of a given event in members of this last population may be unequal to the frequency of the same event in the population considered in the theory of inverse probability.

I hope this will do something to clear up this knotty problem.

M. Fréchet to Fisher: 5 *March* 1940

Thanks for your interesting letter of Feb 26. We are near a point where you will understand in what consists my difficulty and I will be able to see exactly how you come to your conclusion. But not yet quite.

I fully appreciate the difference which you establish between inverse and fiducial probabilities—or at least I appreciate that there is between both an outstanding difference. What remains obscure for me is why you equalize, as a matter of course, two probabilities related to the same event (a given inequality) but computed in 2 different populations. You would be ready to admit this equality as a new principle. This would, in a way, get rid of my difficulty if this new principle would mean an assertion which might be admitted or denied by people who all accept the usual axioms of Calcul. of Prob. But as you consider this new principle as 'the thinking of a given situation in an unfamiliar way', I understand that this is not at all what I meant. Is it right to say that according to your explanation, the considered equalization is an unfamiliar but real consequence of the ordinary rules of Cal. of Probab.? If it were right I think it would be worth the trouble to show how it is a consequence.

You consider also an alternative. The fiducial probability as you have defined it would be an arbitrary definition and I quite admit that it is an 'arbitrariness' (?) limited by rational grounds. But does it mean that after the definition has been given we are not entitled to treat this f. probability as the other probab. (to say that it is defined in a specified population, that it obeys the theorems of total and compound probabilities and so on ?). If it means that, it would be necessary to complete the definition so as to be able to derive something of it in practice. If it does not mean that, if it is a probability like the other ones, a probab. of a given inequality in a given population and if we say that it is equal to the prob. of the same ineq. in a second given population, there is no more a new definition, but a new theorem which has to be proved starting from the ordinary rules of Cal. of Prob. or which has to be admitted as independent of these rules.

You are right in qualifying this problem as rather knotty but I am afraid that many statisticians will use wrongly fiducial probab. as long as all statisticians have not succeeded in agreeing about its real meaning.

Fisher to M. Fréchet: 8 *March* 1940

It was a pleasure to read your interesting letter of the 5th March. You there open up problems, the discussion of which would carry us far, and which I cannot easily attempt by way of correspondence. If someone would supply me with a list of the 'usual axioms of the calculus of probability', or the 'ordinary rules of the calculus of probabilities', I think you or I could, without much difficulty, ascertain whether the probability statements of the kind I call fiducial can be derived from them by a rigorously deductive process.

Now, I do not possess any such list, for an interesting reason which applies to much more of mathematics than our particular problem. Mathematics is the oldest discipline of the human mind, and mathematical truths seem to be the most enduring kind which the human mind can discover. They are not only as solid, but as precious as adamant, but the formulation of the axiomatic

bases from which these truths might be derived is beset with difficulties which we cannot ignore, if only for the reason that the axioms have to be reformulated so frequently. Our palace of adamant rests upon foundations of gossamer which have to be renewed two or three times a week by the indefatigible labours of mathematical logicians, and yet the superstructure seems to be secure and quite habitable.

Certainly one can show, and easily, that fiducial probability, defined as I previously explained, satisfies the laws of multiplication and addition, in fact that it represents, in a well defined population, the proportion of events which belong to a well defined class.

M. Fréchet to Fisher: 11 *March* 1940

We are quite in agreement concerning the difficulty of ascertaining which are the true bases of each science. Though this last problem interests me*, I was simply thinking in my last letter of naive proofs as are given in the many papers on probability. Such a naive proof would quite satisfy me, if using the same principles as those which enable [one] to prove that

$$\{\text{Probability of } (\bar{x}-\mu)/s < t/\sqrt{n} \text{ when } \mu \text{ has a given value}\} = A\int_{-\infty}^{t}(1+z^2)^{-n/2}dz,$$

it were similarly proved the different assertion that

$$\{\text{Prob. } (\bar{x}-\mu)/s < t\,[/\sqrt{n}] \text{ when } \mu \text{ is not given}\} = A\int_{-\infty}^{t}(1+z^2)^{-n/2}dz.$$

As you write that this last equality enables [one] to find the probability distribution of μ, I expect that perhaps the sentence 'μ is not given' means that μ is (like $x_1, \ldots x_n$) a random variable in the second equality.

[P.S.] *I will send you a reprint of my lecture in English at one Cambridge congress 2 years ago on the foundations of the Calc. of Prob.

Fisher to M. Fréchet: 18 *March* 1940

I think the naive proof that you want could run as follows: what we choose to call an 'event' consists of n values x drawn at random from a normal population. The mean of the population we shall designate by μ, the mean of the sample by $\bar{x}$, and the estimated variance by

$$s^2 = S(x-\bar{x})^2/(n-1).$$

These events may be divided into two classes, (a) the successful events for which $\bar{x}-\mu$ exceeds $st/\sqrt{n}$, and [(b)] the unsuccessful events in which it is equal to or less than $st/\sqrt{n}$. Considering the population of all events that occur, without selection, the probability of success is the 'Student' integral from t to infinity, e.g. t can be chosen so the probability of success is 5%, 2%, 1%, etc. So far this is pure mathematics, and, I believe, unexceptionable, Now, suppose we are confronted with a concrete sample for which we accept the belief that it has been drawn from a normal population of which nothing is known except what the sample tells us of the parameters μ or σ; we say 'here is an event which *may be legitimately regarded* as one chosen at random from the population of which the theory was investigated above'. The probability that it is a success is 2%. If this is so, the value of μ is less than a quantity which I can calculate from the sample. In this sense, therefore, the probability that μ is less than this quantity is 2%. The same argument applied to other percentile values gives a consistent series of values of μ, i.e. one in which μ increases when the percentage is increased. The statement that the event with which we are confronted is one chosen at random from all such events is therefore made a basis from which we deduce a frequency distribution for μ, and which may be called its fiducial frequency distribution. In relation to its frequency distribution μ is, properly speaking, a random variable whatever may be the physical origin of the sample considered.

The stipulation that nothing beyond the sample is known of the population sampled is relevant to our decision to regard the concrete sample as one chosen at random from the population of all such samples. It is, of course, only in the sense implied by that decision that the distribution of μ exists.

M. Fréchet to Fisher: 13 *April* 1940

Please excuse my delay in answering your letter of March 18th. . . .

I read your letter with great interest. However, I expected some detail not given in your publications and which would answer the question which I had put out at the very beginning. This question was and remains unanswered: how is it that in your proof, you equalize without comment, as a matter of course, the probabil. of one same inequality in two different populations?

I will try in §I to point out the point in your letter where this arises. In §II and III, I will raise two other questions which were always in my mind, but which I had formerly not stated or at least not clearly stated.

I. Your proof has 2 parts. The first one is, as you say, unexceptionable; but to prepare what I have to say on the second part, I would précise [*sic*] that your probability of success refers to a population where μ (though arbitrary) is given and fixed.

I should think that we both agree,—since you mention that the first part is unexceptionable,—that it is the second part which is difficult.

As a secondary remark, I would say first, that when you write

> 'here is an event which *may be legitimately regarded* as one chosen at random from the population of which the theory was investigated above'

I would replace 'from the population of which' by: 'from one of the populations, the theory of each of which'. But the capital point is in the following 3 sentences. There, to be quite clear, it is necessary to explicit [*sic*] each time the population which is spoken of, though it would be pedantic to do so in one of the usual reasonings on other matters. I insert in your text my addition: 'The probability that it is a success is 2%' (it refers here to the case where μ is *fixed*, though perhaps unknown, if it can be considered as proved). 'If this is so, the (fixed) value of μ is less than a quantity which I can calculate from the sample (which is random, so that the probability is computed before the sample has been drawn, otherwise it would be 0 or 1). In this sense therefore the probability that μ is less than this quantity is 2%' (provided μ is still fixed (though unknown) and the quantity be still random). 'The same argument . . . from which we deduce a frequency distribution for μ . . . .' Then we suddenly pass from the case where μ was fixed to the case where μ is random and we speak of *the* probability of $\bar{x}-\mu > st_0/\sqrt{n}$ (t_0 corresponding in 'Student's' distribution to 2%) as being 2% as well when μ was fixed—the case when 'Student's' dist. is proved—as when μ is random: a different case where 'Student's' distribution *has to be* proved.

II. The conception of the new population which is introduced, formed by all possible normal populations with arbitrary μ and σ, (the drawing of all elements of one sample being done in one (random) of these populations) is not an easy one; and the fact that, nothing being said about the way of choosing at random one of these populations (before drawing a sample in it), there is, however, a definite distribution of μ, may look strange.

III. When a proposition becomes wrong when one of its hypotheses is abandoned, it is always possible to find exactly the point of the proof where this hypothesis had to be assumed.

Now, I understand that you consider that your proof would be wrong if the sample would enter in the proof through inefficient statistics. In fact, you explain in your Paris lecture at the Soc. de Biotypologie [*CP* 156], that the *result* of the proof could not safely hold in that case; but you do not show where, *in the reasoning* itself, the efficiency of $\bar{x}$ and s enters. In no place of the proof quoted above, efficiency or information were mentioned, so that the proof applies or appears to apply to both cases.

The reader may get the impression that you formed your proof first, that you recognized after, that the result of the proof was not valid when inefficient statistics are utilized and that instead of revising the proof, you introduced a new hypothesis. I cannot say that this was your procedure, but that it is how your procedure looks to have been. Perhaps precisely in making explicit the point where this hypothesis appears you may let disappear the gap, or what appears to me the gap, mentioned in §I of this letter.

I wonder whether I have mentioned that by informing me of the existence of Prof. L. Jones' laboratory in Cambridge you increased considerably the value of my enquiry in England. Prof. L. Jones gave me very useful information.

Fisher to M. Fréchet: 18 *April* 1940

I have your letter of April 13th and mine of March 18th before me. As I stated the case in my letter, I do not think it is correct to say that in my proof I 'equalise without comment, as a matter of course, the probability of one same inequality in two different populations'. The only population which I consider in that proof is that of events as defined in the first paragraph.

The question I proceed to discuss is whether, when confronted with a concrete sample, we may legitimately regard it as an event chosen at random from that population. In that population I ought to insist that it is indifferent whether μ is the same from sample to sample, or varies from sample to sample, that is, from event to event, for samples from populations having any value μ are, in fact, members of the population of events defined. I therefore demur at your comment that the value of μ, though arbitrary, is given and fixed. I go on to suggest that, when we have a concrete sample for which we accept the belief that it has been drawn from a normal population of which nothing is known except what the sample tells us of the parameters μ and σ we may legitimately regard it as an event drawn at random from the population investigated. This step seems to me strictly analogous to that which is made in all applications of the theory of probability; for example, when a man whose personal and family history in no way differentiates him from the actuarial population accepted as healthy lives is accepted by a Life Assurance company at its standard rates. This is certainly an act of judgment, and not a deduction from any axiomatic basis, or at least from none that I should be prepared to put forward. Your amendment involving the words 'from one of the populations, the theory of each of which has been investigated' does not seem to me necessary, if the full extent of the original population is once grasped.

If I am not mistaken as to your meaning, this reservation applies to the other difficulties which you feel, which turn, I think, always on the idea that the population of events defined refers to drawings from normal populations having the same mean. The distribution of t is certainly independent of any possible variation in the parameters μ and σ.

Under section III of your letter you raise a distinct point, which I have discussed in some of my addresses on the subject, namely, why I reject analogous arguments based on inefficient statistics. This point is fundamental, and I think quite easily explained. It is that, whereas in deductive reasoning we may make any selection we please from our axiomatic basis, and, reasoning from the selected axioms only, it may be possible to derive certain rigorously justifiable consequences, it is a characteristic of inductive reasoning that the whole of the information available must be utilised. Every statistician is aware that by an arbitrary selection from his data, and the subsequent use of this selected portion as though it were the whole of the information available, he could make a show of justifying any number of false conclusions. The use of inefficient statistics appears to be indistinguishable from a selection of part out of the whole of the data available (see The logic of inductive inference. *J.R.S.S. 98*, 39–82 [*CP* 124]). Of course there is also a sense, elaborated by Neyman and his colleagues, in which formal inferences may be drawn from inefficient statistics, but such formal inferences when properly understood, or when fully stated, are seen to be irrelevant to the objects of scientific research.

M. Fréchet to Fisher: 24 *April* 1940

I will first reply to your last letter.

But there is no hurry to read this reply; whereas I would be much obliged if you could look at my P.S. appended on a *separate sheet* and either reply to this P.S. or send to me the *Metron* paper mentioned there, the soonest being the best.

Now to the thorny question where it seems to me that we are marking time both sides.

I. Concerning my observation about inefficient statistics, you repeat what I read in your printed papers, that is, a valuable *explanation* of how the *result* could not be right with ineff. statistics. But in my letter I warned that I had read this explanation and that what I would like to get was totally different, that was: in which part of the *proof* (composed of 2 parts) of the distribution of μ is introduced the hypothesis that the statistics are efficient?

II. I cannot agree with the statement that you only consider one population; and perhaps our difference comes from the fact that you think that I do not accept what I call your second population. Though, I had not first understood exactly what was this second population, when I thought that μ was still given, you have explained to me your view, substantiated by your paper on inverse probab.

Therefore, I do *not* reject the population mentioned in your last letter, where μ is random, when you come to the *second* part of your proof concluding with the distribution of μ. But, I understand that in the first part of the proof, the one which is unexceptionable, μ is given and fixed and this is where I find a first population $\neq$ [*sic*] from the second though the 2 proba. refer to the same event (the same inequality).

Perhaps one day we may clear the matter, and then I know very well that it will be useful to several statisticians and mathematicians beside myself.

P.S. You were kind enough to indicate to me the solution of the following question (avoiding to me the necessity of computing some multiple integrals already computed):

Let $x_1, \ldots x_n$ [denote] random values independently and normally distributed with some law, $s^2 = S(x_i - \bar{x})^2/(n-1)$, $\bar{x} = Sx_i/n$.

To find the distribution of

$$w = S\gamma_i(x_i - \bar{x})/s = S\gamma_i\, x_i/s,$$

where γ_i are constants such that $S\gamma_i = 0$.

You answered on January 17th that w has the distribution of $t\sqrt{S\gamma_i^2}$ where t has the 'Student's' distribution. Applying this result my assistant found some unexpected consequences and trying to find where was the difficulty, he noticed that the 'Student' t may vary from $-\infty$ to $+\infty$, whereas by Lagrange Inequality:

$$|S\gamma_i(x_i-\bar{x})/s|^2 \leqq S\gamma_i^2 S(x_i-\bar{x})^2/s^2 = (n-1)S\gamma_i^2$$

so that $|w|$ is always $\leqq \sqrt{n-1}\cdot\sqrt{S\gamma_i^2}$, from which it appears that there is something wrong somewhere.

I expect that it is due probably to some misunderstanding somewhere and easy to settle.

As you mentioned that your result can be found in one of your papers in *Metron* 1926, Vol. V, pp. 90–104, and as *Metron* is not easily accessible here, I should be much obliged if you were kind enough to send me (or lend to me) your *Metron* reprint.

(I have not forgotten that you expect me to send you back your paper on Inverse Prob., when I have no more to utilize it. Unless you require it sooner, I will keep it for some time).

Fisher to M. Fréchet: 26th *April* 1940

Looking at the P.S. of your letter of April 24th in order to give you a quick reply, I see that the paradox is partly my own fault. I had sent[1] the solution of the distribution of:

$$S(Yx)/s$$

where x is distributed normally about zero, without noticing that your problem with $S(Y) = 0$ introduces a restriction which diminishes the degrees of freedom by one. For the common form of the analysis of variance we then have

	Sum of Squares	Mean Square
1	$S^2(Yx)/S(Y^2)$	s'^2t^2
$n-2$	$S(x-\bar{x})^2-S^2(Yx)/S(Y^2)$	s'^2
$n-1$	$S(x-\bar{x})^2$	s^2

so that the ratio you enquire about may be equated to

$$S(Yx)/\{s\sqrt{S(Y^2)}\} = t\sqrt{n-1}/\sqrt{n-2+t^2}$$

the distribution of which is easily derived from that of t, being that of the sine instead of the tangent of an arbitrary angle. t has now, of course, $n-2$ degrees of freedom.

[1] See Fisher's letter of 17 January 1940 (p. 119).

M. Fréchet to Fisher: 14 *June* 1947

I send you separately a typewritten copy of my report on the 'Estimation of parameters' about the enquiry organized on that subject by the Intern. Inst. of Statistics. This report shall be presented at the September Session in Washington of this Institute and followed by discussion.

As I tried, in pages 35–36, to explain your position, I think best to send this copy to you, because if your answer would change my mind, I might still introduce corrections when I receive the proof sheets.

Furthermore, I wonder whether you are aware of criticisms of fiducial probability which have been published by such eminent scientists as Gini and Serge Bernstein and von Mises. I am quite convinced that their reasonings are rigorous and that the difficulties lie in what they have understood that fiducial probability is.

As even those who think they closely follow you have been mistaken (you remember perhaps those few lines of Deming and Birge which you repudiated[1]), I think it would be most important if you would scrutinize the papers by these three colleagues (of so high standing) and show where the discrepancy happens. . . .

[1] See Fréchet's letter of 8 January 1940 (p. 118).

Fisher to M. Fréchet: 21 *June* 1947

Thank you for your letter and courtesy in sending me your critical commentary. I shall not be inclined to argue the matter, and no doubt the great statisticians whom you mention will discover in time how the matter should properly be expressed. It seems to me obvious that serious mental obstacles must always have existed to the apprehension of a form of reasoning which seems to me new and valuable, and that they should exist in illustrious minds trained in earlier ideas is not at all to me surprising. I should be glad if anyone reading my works should take what good they may find in them and make good use of it, and trouble themselves little about what they think to be false or defective.

M. Fréchet to Fisher: 18 *October* 1951

Vous savez que j'ai la plus grande admiration pour l'ensemble de votre oeuvres. Vous savez aussi que je ne suis pas toujours d'accord avec vous sur les détails.

Ce sont sur quelques-uns de ces points que j'ai attiré l'attention dans le Rapport au Congrès de Washington de 1947.

Je préfère vous les indiquer moi-même au moment où ce rapport vient de paraître, plutôt que de vous laisser les apprendre indirectement et beaucoup plus tard.

Vous avez sans doute reçu comme moi ces jours-ci le volume contenant ces rapports 'Vol. III, Part A, *Proc. Intern. Stat. Conf.*, 1947, Washington'. Les passages où des observations vous concernent sont aux pages 373, 4e ligne; p. 380, dernière ligne; p. 381; p. 378.

Certains auteurs consultés dans mon enquête, font des objections qui vous concernent, en particulier, von Mises: p. 390.

J'ai tenu à vous mettre au courant, mais bien entendu, il ne m'appartient pas de dire s'il convient ou non que vous répondiez à ces observations.

Fisher to M. Fréchet: 23 *October* 1951

Thank you for your considerate letter of October 18th.

I am not much tempted to enter into controversy, since I have been much impressed now for many years with the number of highly intelligent men in mathematical departments, masters for the most part of an exquisite mathematical facility, who have blundered quite dreadfully in trying to rehandle the mathematical methods which have a purpose to serve in Statistics. Where I have been consulted in time, I have been able on many occasions to dissuade authors from making extravagant claims under a misapprehension of the nature of the problem, but this has not always been possible, and I have come to the conclusion that, given time, and zealous emulation, most of the errors committed in ignorance will clear themselves up, though in a language difficult for the applied statistician to understand.

M. Fréchet to Fisher: 3 *November* 1951

Vous admirez Laplace et pourtant vous n'êtes pas toujours d'accord avec ses écrits.

C'est dans le même esprit que tout en admirant l'ensemble de votre oeuvre et constatant combien vous avez fait progresser la statistique mathématique, je ne puis *parfois* vous suivre dans vos conclusions.

Voici un point que je soulève dans mon rapport et sur lequel je *crois que vous me donnerez raison*. C'est un point *d'histoire*.

Si j'ai bien compris la page 528 de votre mémoire: Inverse Probab. (*Proc. Cambridge Phil. Soc.*, vol. XXVI), vous avez été mal informé. Il vous suffira de lire la citation de Laplace ci-jointe pour voir que Laplace n'a pas ignoré, tout au contraire, le cas des probabilités a priori *inégales*. J'ai en outre, dans mon rapport, donné d'autres références sur ce sujet, p. 372–373.

Il reste un point sur cette question, qui peut être débattu. D'après von Mises et d'après Molina, Bayes lui-même a ignoré ce cas. Il reste possible que Bayes l'ait écrit dans un passage qui leur aurait échappé. Peut-être pourriez-vous le faire connaître? Jusqu'à ce que cette preuve soit donnée, je propose donc d'appeler la formule complète, formule de Bayes–Laplace.

Dans votre dernierè lettre, vous dites que des mathématiciens ont commis des 'blunders' en s'occupant de statistique mathématique. *Je suis tout prêt à l'admettre*, et je crois qu' un mathématicien qui commence à s'occuper de statistique mathématique a tout intérêt à consulter des 'applied statisticians'. Toutefois, *il y a réciprocité* et les erreurs, même non mathématiques, ne sont pas toujours du même côté. Vous vous souvenez peut-être que je vous avais signalé un passage d'un mémoire, d'ailleurs, en général, bien fait, par Birge et Deming et que vous aviez été *d'accord avec moi* pour reconnaître que ce passage renfermait une dangereuse erreur d'interprétation. Pourtant Deming est un 'foremost' 'applied statistician'.

En résumé, et je crois que sur ce point aussi vous serez d'accord avec moi, statisticiens et mathématiciens ont intérêt à se contrôler et à s'aider mutuellement.

[Enclosure]

On a parfois écrit que Laplace avait réduit la formule de Bayes au cas où les probabilités sont égales.

Monsieur Itard signale dans *l'Essai Philosophique sur les Probabilités* de Laplace, 6ème édition, 1840, page 264, le passage suivant, où nous soulignons une phrase, qui suggère une conclusion exactement opposée: 'Bayes dans les *Transactions philosophiques* de l'année 1763, a cherché directement la probabilité que les possibilités indiquées par des expériences déjà faites sont comprises dans des limites données; et il y est parvenu, d'une manière fine et très ingénieuse, quoique, un peu embarrassée. Cet objet se rattache à la théorie de la probabilité des causes et des évènements futurs, conclue des évènements observés; théorie dont j'exposai, quelques années après, les principes, *avec la remarque de l'influence des inégalités qui peuvent exister entre les chances que l'on suppose égales.*'

Fisher to M. Fréchet: 19 *November* 1951

. . . I still feel somewhat strongly that, as I said at that time, 'it is not to be lightly supposed that men of the mental calibre of Laplace and Gauss . . . could fall into error on a question of prime theoretical importance without an uncommonly good reason' [*CP* 84]. The reason, to which later in this paper I ascribe the contradictions which are historically unmistakeable, is the assumption that uncertain inference of all kinds, irrespective of the logical situation in which it is attempted, can be adequately expressed in terms of the single concept of mathematical probability.

I should submit for your consideration now, that if it were indeed true that this single concept were adequate for all purposes we should confidently expect that definitions of probability should have become more exact and better understood with the progress and study of this subject. In a recent and not unintelligent book, however, (*Probability and the Weighing of Evidence*, by I.J. Good) five very distinct meanings of the word 'probability' are found necessary for the discussion.

The procedure which I have preferred in face of a situation of this kind is to choose the oldest clear and useful definition that I could find, namely that of

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.

Click here for next section