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. . . .

L.J. Savage to Fisher: 10 *August* 1952

Thank you for your recent letter.

I hear that you are expected back in Cambridge on the 20th. I shall be here until the 25th, so there is some possibility that I shall have the pleasure of seeing you. My immediate purpose is to mention two technical matters I hope will interest you and about which I would like to hear your reaction.

The first concerns your 'problem of the Nile' (p. 27.257 in *Contributions* [*CP* 137]).

Suppose the land to be divided is represented by $0 \leqq x < \infty$ and that the yield of a plot E is given by $\int_E e^{-\lambda x}dx$, where λ is the reciprocal of the height of flood. I say there can be no equitable plot. Suppose indeed that E were equitable and that $f(x)$ is 0 or 1 according as x is or is not in E, then

$$\int_E e^{-\lambda x}dx = \int_0^\infty f(x)e^{-\lambda x}dx = \beta \int_0^\infty e^{-\lambda x}\,dx \text{ for some constant } \beta,\ 0 < \beta < 1.$$

Therefore, $\int_0^\infty \{f(x) - \beta\}e^{-\lambda x}dx = 0$

for all $\lambda > 0$, or (if you like) all λ in some interval.

It then follows, according to a very well known theorem, that $f(x) = \beta$ for all x, which is absurd.

Doesn't this example show that the 'problem of the Nile' does not always have a solution, or do I mistake the terms of the problem?

Second, I would like to turn to the method of calculating, or may I say defining, fiducial probability, given in the penultimate paragraph of your [*CMS*] page 25.395 [*CP* 135].

Suppose x and y are sufficient for α and β and distributed according to the density $\phi(x,y;\ \alpha, \beta)$.

Let $F(x_0;\ \alpha,\beta) = P(x \leqq x_0|\alpha,\beta)$ and $G(x_0,y_0;\ \alpha,\beta) = P(y \leqq y_0|x_0,\alpha,\beta)$.

F and G are then uniformly and independently distributed from 0 to 1, so your process can be applied; and under favorable conditions it will lead to a probability density $\psi(\alpha,\beta;\ x,y)d\alpha d\beta$. If however the rôles of x and y are reversed, the process may lead to a *different* ψ. I shall give an example that, if I do not miscalculate, illustrates this danger.

Suppose one observes x and y subject to the density:

$$\phi(x,y;\ \alpha, \beta)dxdy = \frac{\alpha^2\beta^2}{\alpha+\beta}(x+y)e^{-(\alpha x+\beta y)}$$

for $x,y;\ \alpha,\beta$ all non-negative.

Clearly x,y are sufficient for α and β.

$$\int_0^y \phi(x,\eta;\ \alpha,\beta)d\eta = \frac{\alpha^2}{\alpha+\beta}\,e^{-\alpha x}\{(1+\beta x) - [1+\beta(x+y)]e^{-\beta y}\}$$

$$\int_0^\infty \phi(x,\eta;\ \alpha,\beta)d\eta = \frac{\alpha^2}{\alpha+\beta}\,e^{-\alpha x}(1+\beta x)$$

$$G(x_0,y_0;\ \alpha,\beta) = P(y \leqq y_0|x_0,\alpha,\beta) = \left\{1 - \left(1 + \frac{\beta y}{1+\beta x}\right)e^{-\beta y}\right\}$$

$$F(x_0;\ \alpha,\beta) = P(x \leqq x_0|\ \alpha,\beta) = \left\{1 - \left(1 + \frac{\alpha\beta x}{\alpha+\beta}\right)e^{-\alpha x}\right\}$$

$$dFdG = \psi(\alpha,\beta;x,y)d\alpha d\beta = \left|J\begin{pmatrix} F & G \\ \alpha & \beta\end{pmatrix}\right| d\alpha d\beta$$

$$= \frac{\partial}{\partial\alpha}\left\{\left(1 + \frac{\alpha\beta x}{\alpha+\beta}\right)e^{-\alpha x}\right\}\cdot\frac{\partial}{\partial\beta}\left\{\left(1+\frac{\beta y}{1+\beta x}\right)e^{-\beta y}\right\}d\alpha d\beta$$

$$= \frac{xy\,e^{-(\alpha x+\beta y)}}{(\alpha+\beta)^2(1+\beta x)^2}\cdot\{(\alpha+\beta)^2 + \alpha\beta(\alpha+\beta)x - \beta^2\} \times \{(1+\beta x)^2 + \beta y(1+\beta x) - 1\}d\alpha d\beta.$$

In view of the two forms in which $\psi d\alpha d\beta$ is written, its integral is 1 and it is non-negative; therefore, formally, a probability density. But obviously if the rôles of x and y had been reversed in the calculation,

$$\psi^*(\alpha, \beta; x, y) = \psi(\beta, \alpha; y, x)$$

would have resulted, which is quite different.

Fisher to L.J. Savage: 23 *August* 1952

. . . About the Nile, you propose that fertility should depend apart from lambda, only on one coordinate x. In that case, the households receive plots stretching from $x = 0$ to infinity, of width y proportional to their needs. However, I have probably missed your point. I have not yet understood your second example.

L.J. Savage to Fisher: 23 *August* 1952

Though I expect to have the pleasure of conversing with you tomorrow, it may be useful for me to write down some remarks on the problem of the Nile suggested by your letter of this morning.

Not having taken your parable sufficiently literally, I thought it would be alright to imagine the terrain to be divided as one dimensional. Two-dimensional examples can be constructed, but whether they fall within your meaning, you alone can say.

1. Let $x(y,z)$ be a one-to-one measure-preserving transformation from the full (y,z) plane onto the half line $0 < x < \infty$, and let the value of a plot A in the (y,z) plane be given by

$$\iint_A e^{-\lambda x(y,z)} dydz,$$

where λ is inversely proportional to the height of the flood. The argument I gave in my first letter shows that no A can be equitable (unless it be almost empty or almost the whole plane).

Perhaps you will dislike this as an example because $x(y,z)$ is horribly discontinuous.

2. Again let the terrain be the whole (y,z) plane but now take the value of a plot A as

$$\iint_A e^{-\frac{1}{2}(y^2+z^2) + yh \cos h + zh \sin h} dydz$$

where h is the height of the flood, $0 \leqq h < \infty$.

Suppose, if possible, that A is equitable and worth an αth of the whole terrain, and let $f(y,z) = 1 - \alpha$ or $-\alpha$ according as (y,z) does or does not lie in A. This implies

$$* \qquad \iint_A f(y,z) e^{-\frac{1}{2}(y^2+z^2) + yh \cos h + zh \sin h} dydz = 0$$

for all $h \geqq 0$. Consider any ϕ, $0 \leqq \phi < 2\pi$; let $h = \phi + 2\pi n$ for n a non-negative integer; and compute (*) for this special h, thus:

$$y = y' \cos \phi - z' \sin \phi$$
$$z = y' \sin \phi + z' \cos \phi$$

$$\iint f(y' \cos \phi - z' \sin \phi, y' \sin \theta + z' \cos \phi) e^{-\frac{1}{2}(y'^2+z'^2)} e^{y'(\phi+2\pi n)} dy'dz'.$$

It follows by a uniqueness theorem for Laplace transforms that

$$\int f(y' \cos \phi - z' \sin \phi, y' \sin \phi + z' \cos \phi) e^{-\frac{1}{2}(y'^2+z'^2)} dz' = 0$$

for all ϕ and y', and thence, by Fourier analysis, that $f(y,z)$ is almost everywhere zero. This contradicts the hypothesis that A is equitable at level α.

Perhaps you intend that the value of every plot should be non-decreasing in h.

3. Such monotoneity is achieved if the terrain is restricted to $y^2 + z^2 < 1$ with the value of A given by

$$\iint_A e^{h(2+y \cos h + z \sin h)} dydz$$

4. If you don't like the terrain in (3) you can smoothly transfer it into almost any other, for example into the whole plane.

I hope one or another of these examples will shed some light on the problem.

L.J. Savage to Fisher: 24 *October* 1952

Certain points were left dangling after our correspondence and very pleasant conversation of last summer, and I am writing this in the hope that you will now have the leisure to help me put some of them out of the way.

The paragraph beginning at the bottom of page 4.20 of the tentative manuscript of my book, of which I believe you have a copy, reads thus:

It has been countered, I think by R.A. Fisher, that if experience systematically leaves [?leads] people with opinions originally different to hold a common opinion, then this common opinon, and it only, is the proper subject of scientific probability theory.

Of course if it is an error to couple your name with this line of argument, I want to delete reference to you here. Even if the allusion is correct, it would be much better to have formal reference, and I wonder whether you can supply one.

You may remember my mailing you, when you were in America, what I thought was an example showing that the problem of the Nile cannot always be solved. You found a fault in my example, which clarified the problem for me, and I submitted a second somewhat more complicated one, which you had not yet had time to appraise when I left Cambridge. I wonder whether you have been able to give that second example some attention. If you find it correct and interesting, I think I'd like to publish it briefly, perhaps in the *Annals of Mathematical Statistics*.

Another technical matter you were not in a position to discuss with me when I was in Cambridge, but on which I hope you can make some remarks now, is what seems to me to be a serious ambiguity in the definition of fiducial probability in multi-parametric contexts. I wrote out one example illustrating this ambiguity, and orally called your attention to the thesis of a Mr.

Williams, which brings out the same point with different examples, and is now filed as Ph.D. 1675 in the University Library.[1]

[1] Williams, R.M. (1949). The use of fiducial distributions with special reference to the Behrens–Fisher problem. Ph.D. Thesis, University of Cambridge.

Fisher to L.J. Savage: 30 *October* 1952

Thanks for your letter. On your first point, I think my name has come in to your statement on page 4.20 in error. At first reading, I do not clearly comprehend the nature of the argument, and feel sure that I never originated or used it.

Please do not hesitate to publish your example showing that the problem of the Nile cannot always be solved. This has long been my opinion and for a long while I thought that Sam Wilks had a demonstration in a particular and interesting statistical case, for he has asserted that no test of significance in Behrens' problem fulfilling certain obvious conditions could exist for which the level of significance was independent of the variance ratio of the two populations sampled.[1] I was pleased to see this statement made, as it knocked the bottom out of the objection to Behrens' own solution, an objection that I have always felt to be based on the misapprehension that tests of significance always referred to unconditionally repeated sampling, and therefore without force. I should, therefore, much like to see a particular case clearly demonstrated in print.

I am still not clear about the supposed ambiguity in the definition of fiducial probability. Perhaps you could discuss it in relation to simultaneous estimation of the mean and variance of a normal population.

[1] Wilks, S.S. (1940). On the problem of two samples from normal populations with unequal variances. *Ann. Math. Stat.* 11, 475–6 (abstract).

L.J. Savage to Fisher: 7 *January* 1953

A word on the problem of the Nile. A few months ago, I suggested that the function

$$\phi(x,y|\lambda) = \exp\{-\tfrac{1}{2}(x^2+y^2) + \lambda x \cos\lambda + \lambda y \sin\lambda\}$$

was a counter example to the general solvability of that problem, because it is quite easy to show that there is no region in the plane such that

$$(*) \qquad \textstyle\int\int_R \phi(x,y|\lambda)\, dxdy = \alpha \text{ for all } \lambda \text{ and for } 0 < \alpha < 1.$$

I hold myself responsible for the demonstration of (*), but I do wish you would tell me whether (*), if granted, does represent an instance of what you would call an unsolvable problem of the Nile.

It has just come to my attention that Feller gave what amounts to an n-dimensional example in *Statistical Research Memoirs*, *2*, (1935), 117–125. He gives in fact, for each n, a family of probability densities $p(x|\theta, n+1)$, (to use his own notation), such that

$$n!\textstyle\int\ldots\int_R p(x_1|\theta, n+1)\ldots p(x_n|\theta, n+1)$$

cannot be a constant $\alpha (0 < \alpha < 1)$ for all θ, if R is contained in the 'valley' $0 \leqq x_1 \leqq x_2 \ldots \leqq x_n$.

May I turn also to my puzzle about fiducial probability, putting it in terms you may find more convenient. Let x and y be distributed according to the density

$$\phi(x,y|\alpha,\beta) = \frac{\alpha^2\beta^2}{\alpha+\beta}(x+y)\exp\{-(\alpha x + \beta y)\};\, x, y, \alpha, \beta \geqq 0.$$

It can be shown that, for all α and β, the variables

$$S = \left(1 + \frac{\alpha\beta x}{\alpha+\beta}\right)\exp(-\alpha x)$$

and

$$T = \left(1 + \frac{\beta y}{1+\beta x}\right)\exp(-\beta y)$$

are independently and uniformly distributed between 0 and 1. This is most easily seen, I think, from the fact that

$$T = \int_0^\infty \mathrm{d}w \int_0^x \mathrm{d}v\, \phi(v,w|\alpha,\beta)$$

$$S = \int_0^y \mathrm{d}w\, \phi(x,w|\alpha,\beta) \Big/ \int_0^\infty \mathrm{d}w\, \phi(x,w|\alpha,\beta).$$

Your recommendation near the foot of p. 395 of Paper 25 of *Contributions* [*CP* 135] leads to a definition of the fiducial distribution of α and β. But the puzzlement is that

$$S^* = \left(1 + \frac{\alpha\beta y}{\alpha+\beta}\right)\exp(-\beta y)$$

$$T^* = \left(1 + \frac{\alpha x}{1+\alpha y}\right)\exp(-\alpha x)$$

leads to a different one. Is there some way out of this apparent ambiguity?

Incidentally, John Tukey has exhibited the same sort of thing in connection with the Behrens–Fisher problem. Has he shown his example to you, and if so what do you make of *it*?[1]

[1] See Tukey, J.W. (1957). Some examples with fiducial inference. *Ann. Math. Stat.* **28**, 687–95.

Fisher to L.J. Savage: 13 *January* 1953

Thanks for your letter with the various puzzles you have put into it. As regards your own difficulty of the ambiguity of the fiducial distribution, do you not think you should use one-valued function[s] so that an element of frequency on one representation shall correspond uniquely to one on another? I do not think that there is anything else of importance in the example you send.

L.J. Savage to Fisher: 23 *January* 1953

I am not sure I understand the following sentence from your recent letter: 'As regards your own difficulty of the ambiguity of the fiducial distribution, do you not think you should use one-valued function[s] so that an element of frequency on one representation shall correspond uniquely to one on another?' If the sentence means what I think it does, it must arise from an error in algebra on your part or mine. The functions S and T defined on p. 2 of my letter of January 7, are certainly single-valued functions of all their arguments. Also, according to my calculation, for each value of T there is exactly one β; and for that β, and any S, there is exactly one α; so the inverse mapping from S and T to α and β is single-valued — similarly for x and y.

Quite aside from fiducial probability, I do wish you would let me know whether the remarks I made on the problem of the Nile do in fact solve that problem, in the sense in which you understand it.

I'm afraid my questions bore you, though they concern matters in which you are the initiator, and indeed the only authority, and about which you have in the not remote past displayed great interest. I must also confess that I am haunted by the suspicion that you may share to some extent the unpleasant opinion voiced by Owen on my visit to your laboratory, namely, that any critical point raised by a mathematician is for the purpose of destroying what others have built. I do hope you, and he, will believe that that is not the case in connection with my questions. I think it more than possible that fiducial probability, like so many of your ideas, is a very good idea indeed. But I can scarcely form an opinion, as long as I believe that the only general definition formulated is quite ambiguous. Of course, in connection with the problem of the Nile, my questions are not critical in any sense at all. You have proposed a problem as to whether such-and-such can always be done — in the suspicion that it cannot. I have adduced instances in which I think that the thing required cannot indeed be done. I am simply asking you to confirm whether these are indeed instances, or whether there are aspects of the problem that I do not yet appreciate.

Fisher to L.J. Savage: 28 *January* 1953

I am sorry I did not make myself clear. The function you choose for S expressed in terms of α is such that one and the same value of S corresponds with two different values of α. I am sure that you noticed this for yourself without my having to mention it, and indeed it seems to be the point of the whole trick.

I am sure that George Owen was not criticising mathematicians in general, he being himself a very highly qualified analyst. He may have implied that mathematical ingenuity could be employed either helpfully or unhelpfully.

L.J. Savage to Fisher: 4 *February* 1953

It does seem to me that the function I choose for S expressed in terms of α is such that the same value of S never corresponds with two different values of α (within the range of meaningful possibilities for α, i.e. $\alpha > 0$).

Indeed,

$$S = \left(1 + \frac{\alpha\beta x}{\alpha+\beta}\right)\exp(-\alpha x)$$

$$\frac{\partial S}{\partial \alpha} = \left\{ -x\left(1 + \frac{\alpha\beta x}{\alpha+\beta}\right) + \frac{\beta^2 x}{(\alpha+\beta)^2} \right\} \exp(-\alpha x)$$

$$= \frac{-x}{(\alpha+\beta)^2}\{(\alpha+\beta)^2 - \beta^2 + \alpha\beta x(\alpha+\beta)\}[\exp(-\alpha x)] < 0,$$

since by hypothesis $\alpha, \beta, x > 0$. Thus S steadily decreases with increasing α, which seems to prove my point.

I hope you will find this calculation correct, and will reconsider my example.

Fisher to L.J. Savage: 11 *February* 1953

Do you mean to assure me that the Jacobian relating S and T to α and β is of constant sign over its whole range?

L.J. Savage to Fisher: 19 *February* 1953

I do assure you that, in my calculation, the Jacobian

$$\frac{\partial S}{\partial \alpha}\frac{\partial T}{\partial \beta} - \frac{\partial S}{\partial \beta}\frac{\partial T}{\partial \alpha}$$

is strictly positive for all positive (i.e. all meaningful) α, β, x, y. To facilitate any check you might wish to carry out, I find

$$\frac{\partial T}{\partial \alpha} = 0; \ \frac{\partial S}{\partial \beta} \text{ irrelevant},$$

$$\frac{\partial T}{\partial \beta} = \frac{-y\exp(-\beta y)}{(1+\beta x)^2}\{(1+\beta x)^2 - 1 + \beta y(1+\beta x)\} < 0,$$

$$\frac{\partial S}{\partial \alpha} = \frac{-x\exp(-\alpha x)}{(\alpha+\beta)^2}\{(\alpha+\beta)^2 - \beta^2 + \alpha\beta x(\alpha+\beta)\} < 0.$$ [1]

[1] No reply from Fisher to this letter has been found.

Fisher to L.J. Savage: 11 June 1957

Many thanks for your letter of June 5th mentioning the arrangements you have made with the State College at Ames to enable me to make a joint visit to your respective campuses.

This I shall very much hope to do, and fulfil the engagement as of the dates you mention. It will be nice to see old friends again at both campuses, and I hope you will give a moment's thought to the subject you would like me to discuss; perhaps at Chicago something like 'The validity of tests of significance for composite hypotheses', adverting to the manner in which the doctrines of Neyman and Pearson have led to a misunderstanding of the problem solved by Behrens.

Of course, there have been so many erroneous tests of significance for composite hypotheses developed from the same theories that it is hard to trace them all out, or 'cover the ground' in the way that would be necessary to prevent students and practical workers from falling into some of the many pitfalls that have been dug. I rather suspect the proliferation of non-parametric tests on this issue.

L.J. Savage to Fisher: 26 June 1957

It is good to hear, as definitely as would be possible at this early date, that you will be visiting us at Chicago and Ames in the fall.

The talk on tests of significance that you suggest would be splendid, as would any of the important themes in your new book. For example, a clear statement of what the fiducial argument is and is not would be extremely helpful to all of us, students and faculty alike. Again, a thorough discussion of ancillary statistics would be interesting, but I suspect that we will probably get that in any event, because it is likely to be relevant to whatever you do discuss. In short, talk about what you like, and you will have a good and appreciative audience.

L.J. Savage to Fisher: 3 February 1958

When we first met in Cambridge, I mentioned a certain one-parametric family of distributions for which the problem of the Nile is not solvable. I think that I may never have given you a formal statement or demonstration of the example, and I enclose one now in case it is of interest to you. For my part, I would be glad to hear what the present state of the problem is. Would my example be interesting enough to publish, or are such examples now a dime a dozen? . . .

[Enclosure]

An example of an insolvable problem of the Nile.

On the square $|x| \leqq 1$, $|y| \leqq 1$, there is clearly a family of probability densities of the form

$$p(x,y|\lambda) = \alpha(\lambda)\exp\{x\lambda\cos\lambda + y\lambda\sin\lambda\} \text{ for } \lambda > 0.$$

The existence of a similar-region implies the existence of a function f essentially different from 0 and carried on the unit square such that

$$\iint dx\,dy\,f(x,y)\exp\{(x\cos\theta + y\sin\theta)(\theta + 2\pi n)\} = 0$$

for all θ and all integers n.

The 'distribution' of $x\cos\theta + y\sin\theta$ under the signed measure induced by f is thus 0 for each θ. It follows, to reproduce a well-known argument, that

$$\iint dx\,dy\,f(x,y)\exp\{i\rho(x\cos\theta + y\sin\theta)\} = 0$$

for all real θ and ρ, so $f(x,y)$ is 0, contrary to assumption.

It is of course trivial so to modify the example that the unit square is expanded to the whole plane.

Fisher to L.J. Savage: 7 February 1958

Thanks for your note and the example. I have, as you know, long thought that such examples could be, or had been found. The interest of publication would, I am sure, lie in the technique of the demonstration of the non-existence of a solution, which should therefore bring out conditions of non-existence recognizable, perhaps, beyond the single example.

Since it was put forward in 1936 [*CP* 137] it has, I believe, always been these conditions of solubility which constitute the problem. Naturally I have never been impressed by the characteristically phoney 'solution' offered by J. Neyman,[1] which Pitman[2] recommends to the American Statistical Association.

[1] Neyman, J. (1934). On the two different aspects of the representative method. *J. R. Statist. Soc.* **47**, 558–675.
[2] Pitman, E.J.G. (1957). Statistics and science. *J. Am. Statist. Ass.* **52**, 322–30.

Fisher to F.D. Sheffield: 11 *February* 1955

I am sorry my reference in *The Design of Experiments* was wrong.[1] I should have referred to part of a series of articles on 'experimental error' in the same journal, namely 1929, vol. 6, pp. 5–11. I asked Maskell to what paper of his I should refer for this point, and he must have given me the wrong reference. I remember he was early confident, being less inhibited than some of us mathematicians, of the propriety of making probability statements about the unknowns behind experimental data. It has now been proved completely that this point of view is right, although I believe there are places in the States, particularly Berkeley, much concerned to avoid this conclusion!

[1] Sheffield had questioned the accuracy of the reference to E.J. Maskell's work in introducing fiducial limits given in Section 62 of *DOE*. Maskell, a plant physiologist who had worked with Fisher at Rothamsted, published a series of papers in *Trop. Agric.* **5** and **6** (1928–9) under the title 'Experimental error: a survey of recent advances in statistical method'. In one of these (*Trop. Agric.* **6**, 7) after discussing the use of 'Student's' t in testing the significance of differences between variety means, he wrote, 'In general, however, the exact probability for each individual difference is immaterial: we wish usually to know what differences we may accept and what should be rejected as not established. For this purpose it is convenient to calculate, from the standard error and the table of t, values which may be called '*significant differences*', which would be exceeded by chance only once in twenty trials ($P = 0.05$) or, if we wish for greater certainty, only once in one hundred trials ($P = 0.01$).'

Fisher to W.A. Shewhart: 2 *February* 1940

I have just received your letter of January 16th, with the news that you are sending me a complimentary copy of your monograph on *Statistical method from the viewpoint of quality control*. This I shall be very glad indeed to have, as I had already heard a good deal about it, and, having followed your earlier work on the subject with great interest, it will be good to see how you have developed it further.

There is a technical point of some importance with respect to quality specifications, which has had rather a devious history in the literature, and may have reached you indirectly. Some time ago I studied one of the natural extensions of 'Students's' distribution, namely, the distribution of t as inferred from a finite sample for a numerical value, the true deviate of which is τ in relation to the population sampled, i.e.

$$\bar{x} + st = \mu + \sigma\tau$$

where, for the sake of algebraic simplicity, I have taken the old fashioned estimate

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

Then I showed that the distribution of t for given τ depended on τ and the size of the sample only. I did not publish the solution until 1931, since it involved functions not at that time tabulated, but it appears in the introduction to

Vol. I of the *Mathematical Tables* published by the British Association [*CP* 91], which contains a fine table by Airey of the integrals of the probability integral, there called the *Hh* functions, which are the functions required. The point is that one might wish to give a guarantee, under penalty, that not more than a certain proportion, specified by τ, of a consignment would fail in some well-defined way; then one could use a value t applied to a random sample experimentally tested in advance to find what limit could be safely specified at a given risk, this being determined by the position of t in its known distribution for the value of τ required. Alternatively, of course, one could find the risk corresponding to any t found to be the value corresponding with a specification required by the buyers.

Technically, it is obvious that this distribution, if adequately tabled, would have very important applications in your field. I pointed out its importance to Neyman,[1] some time before he took his post with Pearson, and was surprised to see later a paper in the *Comptes Rendus* of the French Academy of Sciences by a Polish pupil of Neyman's in which my solution was given without indication of its source, accompanied by a totally inadequate table. I questioned Neyman on the circumstances, and he assured me that his assistant's paper had originally referred to my work as the source of the solution, but that the reference had been excluded when the paper was shortened by the editors![2]

Whatever plausibility this tale may have had at the time, it is certain that Neyman has since referred to this solution, for example in his contribution last year to the Geneva Congress on applications of the theory of probability,[3] as one of the important fruits of his own theory of testing hypotheses, and not as a result available years before that theory was heard of.

During last Winter Mr. Stevens in this Department has been going ahead with the preparation of more adequate tables, which I think may interest you considerably when they can be completed. You may be glad to know of this, as also that misleading statements are liable to be made as to the origin of a method, the relevance of which to industrial contracts I had in mind from the first, as perhaps my early correspondence with you may have shown. . . .

[1] See Fisher's letter of 24 October 1932 to Neyman (p.190).
[2] Compare *CP* 181, p. 143.
[3] Also published in *J. R. Statist. Soc.* **105**, 292–327 (1942).

H.F. Smith to Fisher: 6 *August* 1954

I have long been under the impression that when I was at the Galton Laboratory (and not yet able to appreciate the points involved), I saw a short paper of two or three pages by you claiming that for purposes of a regression analysis it would always be valid to treat the independent variable as fixed even although it might have been observed as a random sample of some variate population. The point is frequently cropping up, although not usually of much importance, and I am unable to quote your reference if it exists or to track it down in any bibliography. I have always meant to ask you when we meet but invariably forget. I have asked many other statisticians and none know the reference, although you are frequently quoted as having given the conclusion as a verbal opinion. . . .

Fisher to H.F. Smith: 27 *August* 1954

I was glad to get your letter, and particularly so as you take up an exceedingly important point. Probably the earliest place in which I refer to the matter is in *J.R.S.S.* LXXXV, page 599 [*CP* 20] where I am discussing the rather wider question of the goodness of fit of regression lines. Its application to the sampling distribution of estimates of the coefficients comes on page 609.

I am inclined at the present time to stress the importance of the principle involved because it cuts at the root of the fallacious approach to tests of significance introduced by Neyman and Pearson, and usually expressed in some such phrase as, 'the frequency found in repeated samples from the same population'. This, as the case of the regression coefficient shows rather clearly, would be quite the wrong question to ask. I believe no one has adopted a mistaken test in this case, but on at least two occasions entirely competent mathematical statisticians, namely E.B. Wilson at the Harvard Medical School, and G.A. Barnard at Imperial College London, have derived erroneous levels of significance for the analogous case of the 2 × 2 table where, as you know, I recommend the calculation of levels of significance from a reference set of samples all having the same margins, and not such a set of samples as might have been derived by a repetition of the actual process by which the particular sample under consideration was obtained.

It is noteworthy that both Wilson and Barnard, when the method was challenged, and after a detailed consideration of the logical position, have frankly agreed that they were mistaken, and have taken the trouble so to express themselves in print.

Reliance on 'repeated sampling from the same population' was certainly responsible for Bartlett's criticism of Behrens' test of significance, and to the succession of more or less indefensible attempts that he has made to obtain an alternative solution.

The relevant phrases to which I refer you are on page 599, '. . . we do not in practice ignore the size of the array', and on the previous page, 598, '. . . we have not attempted to eliminate known quantities, given by the sample, but only the unknown quantities which have to be estimated somewhat inexactly'.

I have only realised recently, chiefly owing to discussion with Barnard, to what extent Neyman and Pearson's approach, including this question of 'repeated samples from the same population', is due to their thinking of a test of significance as though it were a kind of acceptance procedure in which the repeated samples have an objective reality, and are not, as they are with tests of significance, constructs of the statistician's imagination.

Fisher to D.A. Sprott: 13 *January* 1962

I have delayed answering your letter, partly by reason of receiving other exciting correspondence. Even from that darkest Continent, N. America!

Do you know the name of H. Kyburg of the Rockefeller Institute 21 N.Y.? His line seems to be abstract symbolic logic, but he has recently caught fire on the fiducial argument and indeed may be exaggerating its importance.

No doubt you already know Donald Fraser your neighbour from Toronto, who has recently been bitten all over by mad Neymannians in California, but seems to be undefeated. I fancy he has answered the question with which you end your letter, in the negative, by eliminating s_1, s_2 and σ_1, σ_2 and obtaining[1]

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

an equation between ρ and r involving three random variables, for w is normally distributed about zero with unit variance. For any given ρ this gives the distribution of r as I gave it in 1915 [*CP* 4], while for any given r it provides the fiducial distribution of ρ, in a variety of new forms, but agreeing with that I gave in 1930 [*CP* 84], and used in my book [*SMSI*] (1956, 59).

I have been telling him that in the multivariate case, starting with the known distribution of ρ_{12} (marginal) he should get the simultaneous distribution of ρ_{13}, ρ_{23} conditional on ρ_{12}, and thence step up to t variables.

Mauldon seems not to have a glimmer, and Quenouille, who is often quite acute, seems to confuse Likelihood with Fiducial Probability in spite of the fact that the Likelihood has no measurable Reference Set, such as probability must have.

Try to cheer Donald up (and Kyburg too). It would be a pity if people of his quality were driven to suicide by the concerted yell of these Yanks.

[1] This result was in fact derived by Sprott. See p. 853 of Fraser, D.A.S. (1963). On the definition of fiducial probability. *Bull. Int. Inst. Statist.* 40, 842–56.

Fisher to D.A. Sprott: 29 *January* 1962

I have been working at your letter of Jan 24, and it cheered me up, whatever this may do to you. However my eyes are so bad that when my pen fails to write I do not notice the difference!

Fraser sent me what I took to be an invitation to spend a year at Toronto, which you might think a step in the right direction. I told him I should love to see the De Lurys again, but he has not followed the matter up, so perhaps there is a financial crisis in the Dept of Maths.

I confess Kyburg's letter gave me quite a new idea of symbolic logicians. His very general idea of probability with non-differentiable symbols sounds like the opposite pole to my own approach, which is to particularize the different types of uncertainty which a rational mind can experience. Yet his letter seemed sincere, and not a bit demented.

If I decipher you aright you say that my (1956) simultaneous distribution of σ_1, σ_2, ρ does not give the correct distribution of $\rho\,\sigma_2/\sigma_1$, the true regression. What do you compare it with? The distribution for a sample with given $S(x-\bar{x})^2$, or what? But I probably put your mind on to the wrong road, by trying to say that only when the measures of scale s, σ have been eliminated, do you get down to that of ρ. Perhaps I am wrong, but sticking to the 5-parameter distribution in my book, which I think is impregnable, I should like to see the procedure of getting it out of a 5 fold continuous group of transformations.

I have the t-dimensional distribution of r_{ij} $\{\frac{1}{2}t(t-1)$ values$\}$ in the handy form[1]

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

in which $\gamma_{ij} = -r_{ij}\,\rho_{ij}^*$, $\rho_{ij}^* = \rho_{ij}/\sqrt{\rho_{ii}\,\rho_{jj}}$ and

$$F = \int_0^\infty\ldots\int_0^\infty (u_1\ldots u_t)^{N-2}\exp -\tfrac{1}{2}(u_1^2 + + -2\gamma_{ij}u_iu_j - -)\,\mathrm{d}u_1\ldots\mathrm{d}u_t.$$

I have it all written out, but wanted to know whether it would invert to a distribution of ρ_{ij}. In principle I accept that it is not axiomatic that such a distribution exists.

Simultaneous exhaustive estimation need not be a sufficient condition. The example I gave in *Sankhyā* last year shows that strictly it is not a necessary condition. By the way I should like your reaction to that example, arising out of Behrens. The case rather surprised me by suggesting that the fiducial argument can work without exhaustive *estimation senso strictu*.

For bivariate ρ, I get

$$\frac{1}{\pi\,(N-3)\,!}\,(1-r^2)^{(N-2)/2}\;(1-\rho^2)^{(N-3)/2}\times$$

$$\int_0^\infty\int_0^\infty (uv)^{N-3}(N-2+\gamma uv)\exp -\tfrac{1}{2}(u^2-2\gamma uv+v^2)\,\mathrm{d}u\mathrm{d}v,\ \gamma = r\rho$$

which has a suggestive little factor under the integral.

On the whole the only proofs satisfying me in this field involved probabilities of inequalities monotonic over a double range of variables. A continuous group is all right if it leads to such inequality statements. If not, I do not see the reasoning.

[1] Compare *CP* 288, p. 4.

Fisher to D.A. Sprott: 13 *February* 1962

I have just seen your letter, but not yet reflected on it, as I thought it better to lend it to Rao for the day or so before I fly to Adelaide.

You must apply for a personal assistant typist at your next University but the bigger ones are not always better. Think of where the Cambridge Stats Dept has gone since Daniels left it!

You probably have considered that the data needed for the regression do not imply that x is Normally distributed, but those for ρ and thence for σ_1, σ_2 certainly do.

In the case of one variate it does not seem possible to ball up the integration of the bivariate μ, σ distribution, though undoubtedly Bartlett about 25 years ago must have thought so, for he reacted vigorously though confusedly to the idea of such a distribution, denying even that you could get $\mu + \sigma$, although the more general form $\mu + \lambda\sigma$ had been published earlier. Claims have been published putting frequently a merely faulty analysis; though errors of logic are commoner.

I like inequalities because one is not tempted to use them in transformations which are not monotonic, while a certain amount may be hidden in a Jacobian.

e.g. If $P(\rho, r)$ is the probability of getting a correlation less than r from a population with correlation ρ, r being an exhaustive estimate, and P monotonic in both r and ρ unconditionally, then there exists a function $r_P(\rho)$ such that

$$Pr\{r<r_P(\rho)\} = P.$$

Equally we can define $\rho_{1-P}(r)$ so that

$$\rho_{1-P}\{r_P(\rho)\} \equiv \rho \text{ identically;}$$

then using this function as an operator,

$$Pr\{\rho_{1-P}(r)<\rho\} = P$$

or

$$Pr\{\rho<\rho_{1-P}(r)\} = 1-P$$

all since $\rho_{1-P}(r)$ is unconditionally monotonic in r and ρ. The reference set for these statements of probability consists of pairs (ρ, r) for all samples of given size from all Normal populations.

This was the first, and I think may fairly be called the typical fiducial argument. I fancy only a small minority fail to see it for a single parameter, and I have long been sure that there are conditions and stipulations required for more than one.

Fisher to D.A. Sprott: 22 *February* 1962

I believe I ought to clear up what I meant by the use of inequalities.

Pivotal functions must be functions of a parameter (only one to each such quantity) and of a set of statistics exhaustive for that parameter.

They serve as a basis for probability statements provided

(i) their sampling distribution is independent of all parameters,

(ii) their value is Monotonic with the parameter in question.

If there is more than one parameter each such function provides a marginal distribution for the parameter used,

e.g. $t = (\mu - \bar{x})/s$, or s/σ for a Normal sample.

For simultaneous distribution you may have

(iii) a set of pivotals each involving one parameter, and jointly a set of statistics exhaustive for all,

(iv) the simultaneous distribution of these pivotals is independent of all parameters,

(v) each is monotonic *uniformly* for all variations of the others,

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

has a standardized Normal distribution and is monotonic in ρ. It therefore supplies the marginal distribution of ρ

but
$$\frac{r_{12}}{\sqrt{1-r_{12}^2}}\,\chi_{N-2} - \frac{\rho_{12}}{\sqrt{1-\rho_{12}^2}}\,\chi_{N-1}$$

and
$$\frac{r_{13}}{\sqrt{1-r_{13}^2}}\,\chi_{N-2} - \frac{\rho_{13}}{\sqrt{1-\rho_{13}^2}}\,\chi_{N-1}$$

are not I suppose independently distributed, or to take the crucial condition their simultaneous distribution is not independent of $\rho_{12}, \rho_{13}, \rho_{23}$.

The last chapter in *Statistical Inference* is an attempt to illustrate the use of marginals to climb higher.

I think some of your pseudo-fiducials and all of Tukey and Savage's examples are the result of ignoring some of these requirements.

Fisher to D.A. Sprott: 8 *March* 1962

I have just seen your letter of Feb 28.

Of course I do not think the range of application of probability statements to parameters has even nearly been explored. The first examples I put forward were met with such mud-headed incomprehension, combined with rather rancorous hostility, that the statistical world was diverted from the fruitful but difficult job of thinking, to the more enjoyable game of 'taking sides'. The fiducial argument has suffered more from its defenders or rather equivocators than from opponents.

I think the conditions I put down, if I put them all down, are demonstrably *sufficient*; probably some are not *necessary*. e.g. If $t_1, \ldots t_r$ are pivotals with a simultaneous distribution independent of $\theta_1, \ldots \theta_r$, parameters each involved in one pivotal, then for any limits α_i, $i = 1, \ldots, r$,

$$Pr\{t_1 < \alpha_1, \ldots, t_r < \alpha_r\}$$

is calculable as a function of $\alpha_1, \ldots \alpha_r$ and of the statistics in $t_1, \ldots t_r$ but not of θ.

If each pivotal is *uniformly* monotonic in its own parameter, this probability statement is *identical* with a simultaneous inequality statement about $\theta_1, \ldots \theta_r$ whenever this set of parameters has a simultaneous distribution depending only on the statistics of the sample. If, and only if, these are exhaustive, it is impossible to find any function of the observations capable of defining a sub-set to which different probability statements apply.

I fancy this argument can be applied to any particular case which fulfils these conditions.

One would like to know (a) on what conditions pivotals with these properties can be found, (b) are there cases in which the simultaneous distribution of t parameters being accessible, many marginals of not more than s parameters are accessible. If you have 100 Mackinaw Indians and take 200 measurements on each, it is obvious that taking head breadth does not diminish your information about head length, which is a t distribution. (It is I suppose obvious to you and me but not to friend Roy who perhaps out of loyalty to Hotelling gives results which imply the opposite). A simultaneous distribution can be put up for the racial means of any 99 characteristics taken simultaneous, but the data give nothing further about the simultaneous distribution of any 100.

I shall be sorry if it proves impossible to get a simultaneous distribution of $\rho_{12}, \rho_{23}, \rho_{31}$ for a number of one-dimensional margins and Rao tells me some two dimensional margins are obtainable.

I mentioned climbing up from the margins to explain the point of my last example in *Statistical Inference*. Parameters can exist in strata, with the corresponding sets of exhaustive statistics.

I wonder if there can be more than 3 strata.

[P.S.] I think Box argues that assumed Prob. *a priori* though erroneous makes very little difference. Edgeworth argued in this way in 1908. My point is then what is this assumption doing in our reasoning? Why not see what can be said without it?

Click here for next section