Statistical Science
1992, Vol. 7, No. 1, 5–12

Review of Statistical Inference and Analysis: Selected Correspondence of R. A. Fisher (Edited by J. H. Bennett)

**1990, Clarendon Press,
Oxford, England. 380 pages, $90.00.**

George A. Barnard

This volume of Fisher's letters, together with a companion volume also edited by J. H. Bennett (containing his letters on natural selection, heredity and eugenics), completed the corpus of Fisher's major writings just in time for the centenary year of his birth. The companion volume prints the paper on "The Centenary of Darwinism" read by Fisher in Adelaide in 1959. Fisher begins:

> The great advantage of celebrations of Centenaries lies in the opportunity they afford to consolidate what has been learnt in a century, and to fix in orderly relation to each other, and to the whole, the diverse movements, some fruitful, some abortive, which confuse the history of current events. A century affords an opportunity of taking a bird's eye view, and of eliminating unjust and erroneous opinions more speedily than would happen in the absence of such a periodic stocktaking.

Fisher's public style, of which this is a fair specimen, is very rich, so that one sometimes needs to read and reread to grasp his full meaning. The style in his Collected Papers is freer, although it can still be somewhat convoluted. He is much more relaxed in these letters, which therefore form an almost indispensable adjunct to the rest of his works. Professor Bennett has classified the letters under the headings Statistical Inference, Statistical Theory and Method, History of Statistics, Teaching of Statistics, History and Philosophy of Science, and Scientists and Scientific Research. Within each heading, the letters are ordered alphabetically by the correspondent's name. The correspondence with a single individual is made easy to follow by the index.

One point that must be kept in mind when reading this book can be seen from the fact that Frank Yates, Fisher's closest collaborator by far, has only 10 letters (some of the most important ones in the book), whereas there are 43 letters under my own name. This is because, for many years, Yates and Fisher lived near each other, whereas I was never in this position. On the other hand, I did see quite a lot of Fisher when he was President of the Royal Statistical Society, and at other times, so that many of the questions raised in the correspondence were settled in conversation.

One sequence of the letters between us has made me see an unforgettable incident in a new perspective. It occurred at the Centenary meeting of the International Statistical Institute (ISI) held in Brussels in September 1958. At the introductory reception, Fisher, his daughter Joan, George Box and I were together, and the discussion turned to Bayes's famous 1763 paper. Fisher and I differed over what then seemed to me a minor point of interpretation. He thereupon became most abusive, accusing me of threatening to set back the subject for years to come. I could only say that I would go next day to the Bibliothèque Nationale to check exactly what it was that Bayes *had* written. I discovered the next day that the Bibliothèque's run of the Philosophical Transactions began only in 1832, soon after Belgium itself began as an independent state. On returning to London, I found that, as usual, Fisher had been right about Bayes. And now, reading our correspondence for the first time in sequence, it is clear to me why Fisher was so annoyed: he had made his point twice already, in earlier letters, and I had failed to notice.

I report this because it may help to explain the tone of the important letters between Tukey and Fisher, and some of the references to Bartlett and

George A. Barnard is Professor Emeritus of Mathematics at the University of Essex, Colchester, Essex, United Kingdom.

Lindley. For much of his life, Fisher saw himself as battling against ignorance and misunderstanding—most of all, perhaps, in his revolutionary work on biological evolution, although also in his struggles with the elder Pearson and later with Neyman and those who, as it seemed to him, came under Neyman's spell. This meant that those mathematicians whose experience of experimental scientific work enabled them to see that there was more to statistical inference than could be formulated in purely mathematical terms were given a special welcome, but woe betide them if they failed, in Fisher's eyes, to live up to their initial promise and seemed to be going over to the enemy.

Another Fisherian characteristic relevant to these outbursts was his habit of often thinking hard about a problem and arriving at the solution, long before he actually wrote his conclusions. Then, when he did write things up, he had to try, sometimes without success, to recall the exact sequence of his ideas. It was well known that he had the central idea embodied in the theory of k-statistics while on a train journey, but when he wrote his paper on the subject, he found himself unable to recover the argument he had used. One suspects that, in this way, he sometimes was led to produce a less than rigorous argument to support a conclusion that he had reached by a more rigorous route. When gaps in his stated reasoning were pointed out, he would react over-defensively. A case in point, discussed later, arose in connection with the Behrens-Fisher problem.

On a lighter note, it may be seen from my correspondence that Fisher would typically answer me by return mail, whereas my response to him could take months. On one occasion only do I recall having to wait weeks for a reply from him. This occurred when there was a vacancy for the Presidency of the ISI. I wrote to Fisher suggesting that he promote the candidacy of Prasanta Mahalanobis, it being time that the ISI had a President from a developing country and especially from one that had made such effective use of statistical methods as India had. After a long interval, I had a rather vague reply from him. At the following session of the ISI, both Mahalanobis and Fisher were elected Honorary Presidents. I later discovered this proposal had the warm support of the then Secretariat of the ISI, at least partly, because they had for some time been wondering how they could ensure that they would not have to work under either of these two, each of whom would have made a brilliant, but possibly wayward, President.

The warm tone of his long correspondence with Jeffreys is a remarkable feature of this book. Like Fisher, Jeffreys made contributions to natural science that earned him high distinction without reference to his statistical contributions. There never was any question that Jeffreys, like Fisher, well understood those special features of scientific inference that differentiated it from the purely mathematical mode of thought. So, although the geophysicist and applied mathematician enjoyed jousting in public with the geneticist and statistician, they were always able to discuss statistical issues in a friendly way and, indeed, to agree to the extent that, in Jeffreys' words (May 11, 1937), "It would only be once in a blue moon that we would disagree about the inference to be drawn in any particular case, and that in the exceptional cases we would both be a bit doubtful." To this Fisher returned, a little later (in *Annals of Eugenics* **8** 151, 1938), with "Dr Jeffreys says I am entitled to use maximum likelihood as a primitive postulate. In this I believe he is right. A worker with more intuitive insight than I might perhaps have recognized that likelihood must play in inductive reasoning a part analogous to that of probability in deductive problems.... It may thus be said as Jeffreys notes, that the likelihood contains the whole of the information supplied by the observations."

They differed in their usage of the word probability and in their views of the usages of the "old masters." Although Jeffreys accepted the need for distinctions of logical types in probability theory—in a way that could be interpreted as recognizing Fisher's distinction between probability and likelihood—this, to some extent, verbal difference persisted between them to the end.

One important issue on which Jeffreys and Fisher agreed, and on which they disagreed with many other statisticians, was the Behrens-Fisher or two-sample problem:

Observations x_i, $i = 1, 2, \ldots, m$, are normally distributed with mean μ and standard deviation $\sigma\sqrt{m}\cos\theta$, and observations y_j, $j = 1, 2, \ldots, n$ are normally distributed with mean $\mu + \delta$ and standard deviation $\sigma\sqrt{n}\sin\theta$. The four parameters λ, δ, σ, θ are all unknown, with ranges of $(-\infty, +\infty) \times (-\infty, +\infty) \times (0, +\infty) \times (0, \pi/2)$. We wish to test whether the data are compatible with the hypothesis $\delta = 0$.

Using the standard sample notation for sample mean and variance, and setting $S_y^2 = (n-1)s_y^2 = \Sigma(y - \bar{y})^2$ and $S_x^2 = (m-1)s_x^2$, the obvious estimate of δ is the difference of sample means, $\bar{y} - \bar{x}$. The variance of this is σ^2. The proportion of this variance contributed by $\bar{y}$ is $\sin^2\theta$, whereas the proportion coming from $\bar{x}$ is $\cos^2\theta$. Because $(\bar{y} - \bar{x} - \delta)/\sigma$ is a standard normal variable, while

$S_y^2/\sigma^2 n \sin^2\theta$ and $S_x^2/\sigma^2 m\cos^2\theta$ are independent χ^2's on $n-1$ and $m-1$ degrees of freedom, it follows that

$$t(\delta,\theta) = \frac{\bar{y}-\bar{x}-\delta}{\sqrt{\left[S_y^2/n\sin^2\theta + S_x^2/m\cos^2\theta\right]/(m+n-2)}}$$

has Student's distribution on $m+n-2$ degrees of freedom. If θ were known, $t(0, \theta)$ would be the appropriate test statistic whose tail area gave the P-value $P(\theta)$ for the hypothesis being tested. In the absence of knowledge of the other parameters, $r = s_y^2/s_x^2$ is sufficient for θ, and $r\cot^2\theta$ is distributed as F with $(n-1, m-1)$ degrees of freedom, independently of $t(\delta, \theta)$. Both Fisher and Jeffreys then argue, in effect, that θ can be eliminated by averaging $P(\theta)$ over the fiducial, or Jeffreys posterior distribution of θ, derived from the "assumption" that $r\cot^2\theta$ retains its $F(n-1, m-1)$ distribution even though the value of r is known. The only difference between Jeffreys and Fisher is that Fisher derives this "assumption" directly from the assumed absence of knowledge of θ and of the other parameters, whereas Jeffreys expresses the absence of knowledge of the parameters λ, $\lambda+\delta$, σ and θ by assuming that their joint *a priori* density element is $d\lambda d(\lambda+\sigma)d\sigma d\theta/\sigma\sin 2\theta$. That this difference was not thought by Fisher to be serious is clear from Fisher's letter of November 4, 1939 in which Fisher writes: "I have just re-read your note on the Behrens-Fisher formula .... I find I can follow your arguments perfectly, and should disagree, if at all, only on terminology, for you use the distinction between tests of significance and estimation differently from the way I do, including in the latter cases in which the answer is not, properly speaking, an estimate. However apart from this and the propriety of using the *a priori* factor 1/2 when a precise null hypothesis is specifically in view, I think your paper enables me to appreciate your point of view a great deal better than I have previously done."

In a 1942 paper on the single sample problem [*Izvestia Akademii Nauk USSR Série Mathematique* **6** 3-32 1942, Footnote 12 (pp. 18-19)], A. N. Kolmogoroff, discussing Fisher's fiducial argument wrote:

> In this case we should point out that a new axiom of the theory of probability is needed, along the following lines: If the conditional probability $P(A \mid \theta_1, \theta_2, \ldots, \theta_s)$ of an event A is equal to ω for all possible values of the parameters $\theta_1, \theta_2, \ldots, \theta_s$, then the unconditional probability of A exists and is equal to ω.
>
> We should point out that such an axiom is not needed when the problem is treated from the classical point of view, since in that case a prior distribution of the parameters is assumed and the result follows from the standard rules of probability theory.

So far as I know, Kolmogoroff did not discuss the Behrens-Fisher problem himself, but this passage suggests that, had he done so, he might have agreed with both Jeffreys and Fisher, recognizing the assumption of an a priori density as an alternative to Fisher's fiducial argument, leading to the same conclusion in problems of this kind.

The first, it seems, to criticize Fisher's solution to the two-sample problem, was Bartlett, in a paper published in 1936 in the Proceedings of the Cambridge Philosophical Society (PCPS). His correspondence with Fisher begins with a 1933 note on his suggested use of the two words "chance" and "probability" toward a resolution of the disagreement between Fisher and Jeffreys as to the meaning of the word "probability." Fisher's reply is rather stiff, but not unfriendly. There follows a sequence of letters starting in 1935 in which it is clear that Bartlett is unwilling to accept Fisher's notion, in the single sample case, that s^2 contains "all the information" about σ^2 when μ is unknown. He points out that it is the "theoretical statistic" $(n-1)s^2 + (\bar{x}-\mu)^2$ that we need to condition in order to eliminate σ^2 from the conditional distribution and thus to satisfy one of Fisher's "criteria of sufficiency." For Fisher, the fact that μ is unknown, so that the "information" in $(\bar{x}-\mu)^2$ is unavailable, was to be given more logical force. In particular, Fisher is prepared to derive the fiducial distribution for σ^2 from s^2 alone. His position in this respect is made more explicit in his last letter to me, dated March 1962, in which he suggests that parameters and their corresponding exhaustive statistics arrange themselve in strata. This was in response to a letter from me, unfortunately lost, in which I had made suggestions along lines I published in 1963 (*J. Roy. Statist. Soc. Ser. B* **25** 111-114).

Fisher's relations with Bartlett could not have been improved by the fact that his reply to Bartlett's paper, submitted to the PCPS, was refused publication, although there is no evidence that Bartlett was consulted in the matter. Fisher resigned from the Society in circumstances clarified in his 1938 correspondence with Jeffreys, as a result of which he and the Society were ultimately reconciled. He published his reply to Bartlett in his *Annals of Eugenics* in 1937. In correspondence subsequent to 1935, Bartlett drew Fisher's attention to the

counterintuitive behavior of the Behrens-Fisher test when the variance estimates were based on few degrees of freedom; but Fisher did not deal with this in his reply. From then on, their correspondence acquired more and more the character of a "dialogue des sourds."

One factor that made it difficult for mathematicians to accept the Fisherian approach to statistical inference was the dominance in pure mathematics at that time of the axiomatic method. Under the influence of Hilbert, using the abstract approach, tremendous advances had been made in the first quarter of the century in functional analysis and algebra. In the mid 1930s when Fisher's ideas were reaching their full development, the idea had not yet been given up that a full axiomatization of mathematical reasoning might be possible.

Fisher's skeptical attitude to axiomatics was expressed in a letter to Jeffreys dated August 8, 1939: "I am beginning to realize that we know much more about the methods we practice in reasoning than about the systems of logical postulates necessary to justify these methods. Consequently, such a question—which sounds urgent and important to purely deductive minds—as: 'Is there any new principle or postulate required in this method?' cuts very little ice with me because I do not know that the principles or postulates required for previous methods have ever been satisfactorily taped. Unless that has been done, and I respect Whitehead and Russell, Keynes, etc. as I respect the leaders of forlorn hopes, the first question asked is really meaningless. After all, may not the recognition of logical rigor be as empirical as the recognition of the three-dimensionality of space?..." That he continued to be skeptical is shown in an ironic part of a letter dated May 29, 1945 to H. Fairfield Smith:

> In Paris recently I... learned that a certain, I believe very learned, Russian named Kolmogoroff has postulated an axiom, which purports to justify axiomatically certain types of argument for which I am responsible. It runs something like this. If the probability of an event dependent on a number of parameters θ_1, $\theta_2, \ldots, \theta_s$ has one and the same value ω for all possible combinations of the parametric values, then the absolute probability of the event exists and is equal to ω. So now when I am in any difficulty I can assert with confidence 'by Kolmogoroff's axiom'...."

It is reported by John Hammersley that, around 1950, he attended a lecture in Oxford after which he asked Fisher whether fiducial probability satisfied Kolmogoroff's axioms. Fisher replied, "What are Kolmogoroff's axioms?" On being told what they were, Fisher refused to answer. One wonders whether he thought the list he had been given was complete.

Fisher was never strongly interested in the controversies surrounding the foundations of mathematics that went on during his lifetime, and he did not live to learn of the effective cul de sac produced by Cohen's work on the consistency of the axiom of choice and of its contrary (*Proc. Natl. Acad. Sci. USA* **50** 1143–1148, 1963; **51** 105–110, 1964). In conversation with me he made clear that the fundamental reason for his skepticism was his clear recognition of the fact that, in statistical inference, there is an important distinction to be drawn between "p is true" and "p is known to be true," whereas in pure mathematics such a distinction is without importance. Axiomatization of an area is of great assistance in teaching it and in settling disputes between mathematicians. In conversation, Fisher encouraged attempts at partial axiomatization of principles involved in statistical inference, but he was keenly aware of the fact that we are a long way off any approach to a full axiomatization of the manifold types of uncertainty that can arise.

The Behrens-Fisher problem exposed very clearly the fundamental differences between Fisher and Jeffreys on the one hand and Neyman and Pearson on the other concerning the interpretations to be put on tests of significance. Although Fisher, as referee, recommended the classic Neyman-Pearson 1933 paper for publication, and the Fisher-Neyman correspondence for 1932–1933, recorded here, is very friendly, it would seem that he failed to perceive the fundamental difference between his own approach and that of Neyman and Pearson to conditioning, especially in the case of what Neyman and Pearson called "composite hypotheses." The problem occurred in the 1920s in RAF's controversy with Karl Pearson over the degrees of freedom in χ^2. In testing for normality of shape of a variate distribution, the unknown location and scale parameters are what Neyman and Pearson called nuisance parameters, for which Fisher saw linear functions of the cell frequencies exist that estimate them exhaustively. The χ^2 test for normality is therefore a conditional one, conditioned on the observed values of these linear functions. Similarly, the test for independence in a 2×2 table is a conditional one. Such a view contrasts with that expressed by Neyman and Pearson, who laid it down (*Joint Statistical Papers*, Cambridge, p. 163) as axiomatic that the advance probability, before the results are to hand, of wrong rejection of the hypothesis being tested, must be equal to the chosen "size" α, whatever the values of the nuisance

parameters. For Neyman and Pearson, this requirement of "similarity," seemed essential to their interpretation of the test's validity in terms of long run frequency of "errors of the first kind." For Fisher, on the other hand, it seemed obvious, with a conditional test, that the conditional probabilities of error were the quantities to be looked at and, in general, these could not be known until the sample values were available.

That the divergence of view was not almost immediately obvious was because of several accidental circumstances. Perhaps the most important of these was the limited computing facilities available in the 1930s and the expense of tabulation of mathematical functions. Although the successive editions of *Statistical Methods for Research Workers* gave tables of several percentage points of Student's t and Karl Pearson's χ^2, the double entry requirement for the variance ratio table meant that, at first, only the 5% and 1% points were given. Although, thanks to the then young Dr. W. Edwards Deming, they were later supplemented by the 0.1% points, p-values of 0.05 and 0.01 came to take on the almost magical status that they continue to hold in some quarters even today. The need to relate the interpretation of the p-value to the sensitivity, given the data, of the test was largely overlooked. So much so, indeed, that when a paper by Welch (*Ann. Math. Statist.* **10** 58–69, 1939) showed that holding the size α fixed in sampling from a rectangular population with unknown location resulted in inefficiency of the conditional test, the result was for a long time taken to mean that conditional tests were inefficient, although the inefficiency really resulted from the fixing of α.

The limitations on computing possibilities in the 1930s also led to undue concentration of attention on sampling from distributions of the exponential family, where the distinction between conditional and unconditional tests often disappears.

Although the conditioning issue can arise when there is no "nuisance parameter," for example in sampling from a rectangular population with unknown location, and in sampling from a bivariate normal population with unknown regression coefficient, its effect becomes much more marked when nuisance parameters are present. Thus, even Student's t-test can be looked at in two ways. For Fisher, the sample variance s^2 is, in the absence of knowledge of the mean, μ, an exhaustive estimate of the nuisance parameter σ^2, and so forms the basis of a fiducial distribution for σ. If σ were known, the p-value $P(\sigma)$ associated with a given deviation $\bar{x} - \mu$ of sample mean from population mean is the tail area of the standard normal deviate $(\bar{x} - \mu)\sqrt{n}/\sigma$. Averaging this over the fiducial distribution of σ gives the tail area of the t-distributed variate $(\bar{x} - \mu)\sqrt{n}/s$. It happens that, in repeated sampling from a population with fixed μ and σ, $(\bar{x} - \mu)\sqrt{n}/s$ also has the t-distribution. The requirement of conditionality and the requirement of similarity both lead to the same test procedure. But this fact is peculiar to the normal distribution. For samples, from a nonnormal distribution, the distribution of $(\bar{x} - \mu)\sqrt{n}/s$ must, in the Fisherian approach, be conditioned on the sample configuration c, the vector with ith component $(x_i - \bar{x})/s$; and if the sample has outliers, as can happen moderately frequently, for example, with a Cauchy distribution, or with a skew distribution, the conditional distribution of $(\bar{x} - \mu)\sqrt{n}/s$ may differ markedly from that of Student's t. Until the 1960s, the absence of computers made it impracticable to derive conditional distributions except in highly artificial cases, so that in practice dubious assumptions of normality had to be made or, when such assumptions were excessively dubious, "robustifications" had to be introduced. With modern computers, we should consider a plausible range of distribution shapes and examine whether, for our particular sample, uncertainty as to distribution shape could seriously affect our inference.

When Bartlett drew attention to the failure of similarity of the Behrens-Fisher solution to the two-sample problem, and it appeared that most statisticians had come to accept what Fisher called an "intrusive axiom," Fisher pointed in vain to the bivariate normal case, when we are interested in the regression coefficient β of y upon x. With n pairs of observations (x, y) with sample means $(\bar{x}, \bar{y})$, the estimate of β is $b = \sum\{y(x - \bar{x})\}/\sum(x - \bar{x})^2$ and the estimated variance of y for given x is $s^2 = \sum(y - Y)^2/(n - 2)$, where $Y = \bar{y} + b(x - \bar{x})$ is the estimated value of y for the observed value x. The test statistic $(b - \beta)/s$ is distributed as $t/\sqrt{\sum(x - \bar{x})^2}$, where t has Student's distribution on $n - 2$ degrees of freedom. The resultant test is obviously more sensitive to given errors $b - \beta$ when $\sum(x - \bar{x})^2$ is large than when it is small, and, in practice, our choice of α should be adjusted in the light of the sensitivity required. However, the undue concentration on 5% and 1% p-values led many people to overlook the need to relate α to the required sensitivity. Somewhat elaborate proofs that the (one-sided) test using a fixed α value, regardless of the sensitivity, was "uniformly most powerful similar" encouraged those who enjoyed mathematical theorems to overlook the serious limitation on the procedure entailed by the final adjective. At the same time it was only much later that Fisher himself did much to disabuse people of the

idea that the two numbers 0.05 and 0.01 could be singled out as special.

The 2×2 table, the discussion of which takes up much of the early correspondence between myself and Fisher, might have clarified the conditionality issue had the discreteness of the distributions in this case not introduced further confusion. The fact that the uniformly most powerful similar test of size 0.05 in this case requires rejection of the null hypothesis, once in twenty times, even when "all the animals die," as Fisher writes, should nonetheless have given food for thought to anyone who insisted on similarity.

Yet another source of confusion has already been referred to in the 1936 Bartlett-Fisher correspondence. With the usual $(\bar{x}, s^2)$ notation for sample mean and variance, Fisher says to Bartlett, "The whole point of my procedure, as I think must be clear in my book, is to retain s^2 as the sole available fact about the precision of the average." As previously noted, Bartlett had adopted as his definition of "sufficient statistic for parameter θ" the requirement that the conditional distribution of sample values, given the sufficient statistic, should not involve the parameter θ. With this definition, s^2 is not sufficient for the variance σ^2; we need to have the "theoretical statistic" $(n-1)s^2 + n(\bar{x}-\mu)^2$. With the definition of sufficiency adopted by Bartlett, conditioning on a statistic that was sufficient for a nuisance parameter would leave a distribution free of that parameter, which could therefore yield a test that was both conditional and similar. Because $(\bar{x}, s^2)$ are jointly sufficient for (μ, σ), in both Bartlett's sense and Fisher's, the usual tests for normality, for example, are both conditional and similar. The same logical point arises in the analysis of dispersion on a sphere, where the sample consists of a set of n unit vectors that might represent the direction of magnetization of a set of rock specimens. Assuming the vectors to follow a "Fisher distribution" with pole $\mathbf{p}$ and dispersion parameter κ, the resultant R together with the angle between R and $\mathbf{p}$ are jointly sufficient for the two parameters. But the length $|R|$ of R, on which Fisher bases his fiducial distribution for κ, does not satisfy Bartlett's definition of sufficiency. The resulting test derived by Fisher is not similar. The Williams-Watson test for this problem uses Bartlett's definition and is similar, but it suffers from certain paradoxical features requiring care in its interpretation (*J. Roy. Statist. Soc. Ser. B* **25** 111–114 1963; *J. Appl. Probab.* **19A** 293–303, 1982).

When Welch produced tables for his "approximately similar" test for the two-sample problem, Fisher was able to point to the anomaly that when two samples appeared to confirm a provisional judgment of equality of variances, Welch's test, which failed to make use of this judgment, appeared to be more sensitive than the uniformly most powerful t test that did make use of it. However, the most severe blow to the concept of similarity was dealt by Linnik after Fisher died. Linnik showed that any exactly similar test satisfying other logical requirements would necessarily reject the null hypothesis with positive probability for arbitrarily small values of the standardized difference $d = (\bar{y} - \bar{x})/\sqrt{s_x^2/m + s_y^2/n}$ between the sample means. Linnik's work was followed by the proof of the result that seems to have been conjectured much earlier by Wilks, that no pivotal quantity capable of forming the basis of a similar test can take more than three values. The correspondence with Wilks on the Behrens-Fisher problem, which is recorded in this present book, is tantalizing in that it approaches the problem here involved but does not quite reach it. The fact that Fisher knew of Wilk's conjecture strongly suggests either that further letters are missing or that Wilks met Fisher in England during World War II. In light of these results, it now appears that the "requirement" of similarity for a test resembles the "requirement" of unbiasedness of an estimate. It is convenient if it happens to be satisfied, but to demand that it be satisfied can lead to absurdity.

The Behrens-Fisher problem arises again in the correspondence with Yates, when Fisher was drafting his *Statistical Methods and Scientific Inference*. Referring to some objectives I had raised, Yates suggests to Fisher that he should explain why one is required to condition on the observed ratio of sample variances in the Behrens-Fisher problem, but one is not required to condition on the observed treatment-plot configuration when, for example, the randomization process has produced a Knut Vik square. Fisher's reply repays careful study.

The last of the letters I received from Fisher, and the set which subsequently passed between Fisher, Sprott, Fraser and Rao, make clear that Fisher was continuing to try to clarify the fiducial argument right up to his untimely death. Difficulties with the bivariate normal distribution had been signaled by Tukey, who pointed out that one could impose a rotatory transformation on the variates of a circular normal distribution of unknown mean (μ_1, μ_2) in such a way that a family of pivotals existed, all apparently satisfying the requirements of the fiducial argument, yet which generated nonequivalent fiducial distributions for (μ_1, μ_2). Both Tukey and Fisher seem to have thought (mistakenly, in my view) that uniqueness of the fiducial distribution, on given data, is a necessary requirement; so that, the nonuniqueness implied in Tukey's examples

presented a serious problem. It is not clear from the letters printed here what was Fisher's technical reply to Tukey. That Fisher failed to get his point, whatever it was, across is clear, and his disappointment with Tukey is obvious from the outrageous language in which it is couched. Tukey's arguments can be dealt with along lines similar to those that deal with another paradox noted by Mauldon. This forms part of the issues discussed between Fraser, Sprott et al. In retrospect, it seems best dealt with in terms of the pivotal model for inference, which differs from the customary formulations in allowing explicitly for the fact that the shapes of the distributions of the variates we deal with cannot be known with exactitude.

Data (x_i, y_i), $i = 1, 2, \ldots, n$ that would ordinarily be taken as independently distributed in a bivariate normal distribution with unknown parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ can be described by a *pivotal model* with basic pivotals:

$$p_i = \frac{x_i - \mu_1}{\sigma_1}, \quad q_i = \frac{y_i - \mu_2}{\sigma_2} + \beta(x_i - \mu_1),$$

with (p_i, q_i) independent and *approximately* normal, and $\beta = \rho\sigma_1/\sigma_2$. They can also be described by a similar model in which the x_i and y_i and the suffices 1 and 2 are interchanged. However, it is only when (p_i, q_i) can be taken as *exactly* normal that the two models are equivalent. If inferential operations are restricted to operations on the basic pivotals, Mauldon's paradox can be avoided. A similar restriction avoids Tukey's difficulty. It seems to me not impossible that, in his discussion with Tukey, Fisher had in mind some such idea, but was unable to formulate it clearly.

Fisher's late correspondence, together with his late annotations to his collected papers, makes it clear that he no longer held to the strong assertions concerning fiducial distributions of parameters that he made in the late 1930s and early 1940s. If Kolmogoroff's new axiom is accepted, or if Fisher's own arguments concerning the meaning of probability are accepted, we are faced with a situation not dealt with by the axioms that Kolmogoroff had formulated in 1933. We may know, for example, that (say) the quantity $(\bar{x} - \mu)\sqrt{n}$ is nearly $N(0, 1)$, and we may then come to know that $n = 16$ and $\bar{x} = 3.5$. If we know very little about μ other than what is provided by this knowledge, what we know about the actual value of $4(3.5 - \mu)$ is very close to what we know about the value of an observation that has been taken from $N(0, 1)$, but which remains unknown to us. It would seem to follow that we should regard μ as nearly $N(3.5, 0.25)$. However, μ is a parameter, not an observable. We may not be able to attach any definite meaning to functions of μ such as μ^2 or $\operatorname{sgn}\mu$ that are not 1-1 functions. This is one property that distinguishes parameters from observables. If we have an experiment in which it is possible to observe the value of x, we can so modify the experiment as to observe only $\operatorname{sgn} x$, and this fact allows us to treat observables as random variables in the sense of Kolmogoroff [i.e., as functions defined on a probability space, (measurable) functions of which are in turn themselves random variables]. Sometimes a parameter θ is to be regarded as merely a label for a class of experiments and, in such a case, no clear meaning can be attached to functions such as μ^2 or $\operatorname{sgn}\theta$. On the other hand, parameters such as λ and σ that appear in a pivotal model involving dimensional observations x_i with the distribution of $p_i = (x_i - \lambda)/\sigma$ specified in some standard form must have the same dimensionality as the x_i. Then, a function of λ and σ, such as $\lambda + 2\sigma$, may be meaningful, whereas a function such as λ^3 may not. The fiducial argument appears to enable us sometimes to make probability statements about parameters without implying that parameters can be treated as Kolmogoroff random variables. This reconstruction of what Fisher may have been thinking is, of course, conjectural, although it is a fascinating exercise to trace the hints in this direction that can be found, for example, in his correspondence with the economist Roy Harrod, who had written a book about induction that received a favorable review from Fisher.

I have concentrated on some of the major themes arising in the correspondence that are relevant to current more or less technical discussions on the foundations. There is, of course, much else to be learned from this book. The correspondence with Darmois, some of which was used by Fisher in his *Statistical Methods and Scientific Inference* will be most helpful to anyone who is baffled by Fisher's 1925 paper on the "Theory of Statistical Estimation." In the correspondence with Savage, the idea, which was current at the time, that Fisher's "Problem of the Nile" had a major role in the general theory of statistical inference is disposed of. The correspondence with Finney on various aspects of the fiducial argument, and on many points that arise in practical applications, is full of good things. The letter to E. B. Wilson, in which Fisher discusses "one-sided" and "two-sided" tests, puts these issues in a way that is most helpful.

Professor Bennett cannot be criticized for omitting the very important correspondence between Fisher and "Student" (W. S. Gosset). Fisher kept Gosset's letters, but nearly all of Fisher's replies appear to have disappeared, perhaps in a "holocaust" occasioned by Gosset's transfer from

Dublin to London in 1935. In 1957 Lance McMullen visited Fisher in Cambridge when Fisher was clearing up preparatory to leaving Cambridge for Australia and McMullen succeeded in rescuing Gosset's letters, with some brief notes on these made by Fisher. When Fisher died, McMullen circulated copies to a number of statisticians so that they are now moderately widely available. They were used and extensively quoted in the book *Student – A Statistical Biography of W. S. Gosset* by Egon Pearson (edited by R. L. Plackett, with the assistance of G. A. Barnard, Oxford University Press, 1990), which thus provides a useful supplement to Professor Bennett's edition. It can be seen, for example, that when Fisher said of Neyman's "intrusive axiom," referred to previously, that it was "foreign to the reasoning on which the tests [of significance–G. B.] were based," he certainly wrote truly so far as Gosset was concerned, because Gosset clearly thought in terms of uniform prior probabilities applied to parameters such as his population means.

The quotation with which I began, in which Fisher gives his views on centenaries, is of doubtful application to Fisher's own centenary. In this age of uncertainties of many different types, it may be doubted whether any finite limit may ever be set to the range of application of statistical ideas of one kind or another or to the range of concepts that are involved. Fisher's work may come to be seen as a central core of concepts familiarity with which it is an essential basis for applications in many diverse fields. In his last letters and papers, he made it clear that a vast territory remains to be explored.