lies near the middle but happens to be a point with low probability function. But if the parameter space consists of a single point, Birnbaum's principle M would require that the two points x and x' would have the same evidential meaning. This suggests that in significance testing situations the likelihood principle fails to apply, which is, of course, what many statisticians think.

4. Minimal experiments

Kalbfleisch's concept of a minimal experiment is highly suggestive. While, perhaps necessarily, it remains a little vague, it points towards a deep-lying notion in the philosophy of natural science in general, and of statistical inference in particular. This is the notion of a repeatable experiment (Fisher, 1966, Chapter 2, §7): 'In order to assert that a natural phenomenon is experimentally demonstrable we need, not an isolated record, but a reliable method of procedure ... etc.)'. What, minimally, must be repeated is a minimal experiment in Kalbfleisch's sense.

5. Applicability of Birnbaum's conclusions

Although these considerations may be taken as ways of avoiding the unpleasant consequence of the likelihood principle, and although the present writer does not subscribe to the general validity of this principle, it should perhaps be said that in those cases where Birnbaum's model really does apply, where there is no penumbra to the parameter space, where there is no ambiguity about the definition of the sample space and where Birnbaum's principle M applies, then the likelihood principle does appear to the present writer to be applicable. The growing practice, among geneticists and high-energy physicists, of reporting the likelihood function or its equivalents, for example, the score function, in cases of this kind, suggests that the principle accords with their intuition also, in such cases. It may be noted that conditionality arguments of the kind discussed in § 1 above are consistent with the likelihood principle; the set of observations $(x_1, ..., x_n)$ and the equivalent single observation x give the same likelihood function, though the sample spaces are different.

References

Barnard, G. A., Jenkins, G. M. & Winsten, C. B. (1962). Likelihood inference and time series. *J. R. Statist. Soc.* A **125**, 321–72.
Birnbaum, A. (1962). On the foundations of statistical inference. *J. Am. Statist. Assoc.* **57**, 269–306.
Fisher, R. A. (1966). *The Design of Experiments*, 8th edition. Edinburgh: Oliver and Boyd.
Kalbfleisch, J. D. (1975). Sufficiency and conditionality. *Biometrika* **62**, 251–9.

[*Received December* 1974]

Comments on paper by J. D. Kalbfleisch

By OLE BARNDORFF-NIELSEN

Department of Theoretical Statistics, Aarhus University

The concept of experimental ancillaries is interesting, but Kalbfleisch (1975) has, in my view, not produced convincing reasons for being unable either to apply S before C_E, or to apply C_M before S. I am inclined to think that such reasons of principle cannot be found and that the resolution of the dilemma raised by Birnbaum's theorem lies, not in prescribing some fixed sequence of modified versions of C and S, but in accepting that statistical analyses, in general, will not and cannot lead to unequivocal conclusions, and, in particular, that various applications, to the same model and data, of the procedures of margining to a sufficient statistic or of conditioning on an ancillary statistic may lead to different views on the chance mechanism.

It is to be expected that reasoning from different angles about a subject matter on the basis of incomplete information will result in differing answers. Ambiguity of this kind is a difficult but not prohibitive feature of scientific investigations. For instance, there is nothing fatally paradoxical about the general lack of nonuniqueness of maximal ancillaries, as has recently been stressed by Barnard (1974).

Now, Birnbaum's result may be paraphrased as saying that if it is required that application of the ideas of sufficiency and conditionality never leads to conflicting, or nonequivalent, conclusions then these conclusions have to obey the likelihood principle. But, on the above viewpoint, it is not reasonable to impose such a requirement, and Birnbaum's theorem may be taken as showing that sufficiency and conditionality do not satisfy the requirement.

Even if Kalbfleisch's proposal to start by applying C_E is not considered cogent, the distinction between experimental and mathematical ancillaries is likely to be useful. On the matter of which ancillaries are to be taken as experimental, I wonder what Kalbfleisch's attitude would be to Example 1 if our knowledge of the random mechanism implied that $f(x)$ had to be of the form (1), or to the first part of Example 2 if, similarly, it was known that the regression was linear with normal errors. Would this knowledge change the ancillaries in question from mathematical to experimental?

It may be illuminating to consider the further example of the two-by-two table, Table 1(*a*), obtained by classifying a random sample of n individuals according to phenotype at two diallelic loci with dominance. Assuming that the population sampled is the offspring of a population consisting entirely of double heterozygotes of trans-type, and that there has been random union of gametes and no selection, the corresponding table of probabilities is that of Table 1(*b*), where the parameter π is the product of the recombination frequencies for males and females.

The statistics $x_{1.}$ and $x_{.1}$ are maximal ancillaries, but are they mathematical or experimental, or is this a case which we have to leave undecided?

Table 1. *Two by two tables*

(*a*) Observations

	$A-$	aa	
$B-$	x_{11}	x_{12}	$x_{1.}$
bb	x_{21}	x_{22}	$x_{2.}$
	$x_{.1}$	$x_{.2}$	n

(*b*) Model

	$A-$	aa	
$B-$	$\frac{1}{4}(2+\pi)$	$\frac{1}{4}(1-\pi)$	$\frac{3}{4}$
bb	$\frac{1}{4}(1-\pi)$	$\frac{1}{4}\pi$	$\frac{1}{4}$
	$\frac{3}{4}$	$\frac{1}{4}$	1

The chance mechanism under study is the recombination process, and it may be argued that the ancillaries have come about through the design of the experiment and that $x_{1.}$ and $x_{.1}$ are therefore experimental ancillaries. Provided $x_{1.}$ and $x_{.1}$ are considered experimental, then we have here a clear cut instance of nonuniqueness of maximal experimental ancillaries, with nothing to choose between $x_{1.}$ and $x_{.1}$. And the indication is that if an unequivocal answer of the statistical analysis is, unreasonably, demanded then one is, in effect, forced to obey the likelihood principle.

It may be noted that the minimal sufficient statistic for the original model is obtained by adding x_{12} and x_{21}, and after this reduction there seems to exist no ancillaries.

REFERENCES

BARNARD, G. A. (1974). The foundation of statistics 1964–1974. *Bull. IMA*, **10**, 344–7.
KALBFLEISCH, J. D. (1975). Sufficiency and conditionality. *Biometrika* **62**, 251–9.

[*Received November* 1974]

Comments on paper by J. D. Kalbfleisch

BY ALLAN BIRNBAUM*

Department of Statistics and Computer Science, University College London

As one who is inclined to use conditional inference with standard statistical methods, relying on partly *ad hoc* considerations of familiarity and simplicity in the face of theoretical puzzles, in the spirit of Cox (1971), I find Kalbfleisch's (1975) proposed modified conditionality concept interesting and potentially useful, but unsatisfactory in respects which will be explained below.

* Present address: Department of Mathematics, The City University, London.