Comments on Paper by J. D. Kalbfleisch

Author(s): Allan Birnbaum

Source: *Biometrika*, Aug., 1975, Vol. 62, No. 2 (Aug., 1975), pp. 262-264

Published by: Oxford University Press on behalf of Biometrika Trust

Stable URL: https://www.jstor.org/stable/2335359

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.

The recent period of interest in the conditionality, or ancillarity, concept may be said to have begun with Fisher's (1956, pp. 156–9) insistence on the importance of that concept for the further adequate development of the theory of statistical inference. He did not anticipate the sorts of 'monsters' which would be encountered along that path; the terms 'monster-barring' and 'monster-adjustment' were used by Lakatos (1963–4) to describe developments in the evolution of the concept of a polyhedron, which occurred in the course of many attempts to prove or refute a famous conjecture of Descartes and Euler.

In earlier writings Fisher had given the mathematical definition of an ancillary statistic; a rule of application, replace unconditional by conditional models before applying specific standard techniques of statistical inference; and an unformalized but suggestive interpretation of conditionality as a theoretical concept. All of these were considered unproblematic, appropriate, and important by many of the theoretical and applied statisticians acquainted with them, including Cox (1958) and J. W. Tukey in unpublished Wald lectures.

Monsters appeared soon after 1956 in connexion with both the applied and the theoretical aspects.

(*a*) The rule of application was seen to be an ambiguous directive when ancillaries were found to be nonunique in general (Basu, 1959, 1964; Birnbaum, 1961).

(*b*) The theoretical concept was found to be essentially equivalent to another concept, likelihood, incompatible with use of standard statistical methods (Birnbaum, 1962, 1972).

The recent contributions of Durbin (1970) and Kalbfleisch (1975) propose successively modified rules of application, intended to systematize appropriately the typical judgements and applications of conditionality and related concepts, in 'the operational approaches of both the Fisherian and the Neyman–Pearson schools of statistical inference' (Kalbfleisch, 1975, §1). These rules eliminate many but not all cases of nonuniqueness, leaving the choice among remaining alternative ancillaries to possible further rules, such as that of Cox (1971), having admittedly a partly *ad hoc* character. These rules are supported by new theoretical interpretations of conditionality and related concepts. We shall discuss here primarily those of Kalbfleisch, which in a sense incorporate those of Durbin. Related discussion of Durbin's paper appears in my 1970 note. The ideas of that note and the present one are also broadly applicable to the discussion of conditionality and related concepts by Cox & Hinkley (1974).

Neither the consequences in applications, nor the theoretical content and form, of Kalbfleisch's modified conditionality concept seem sufficiently clear, particularly in the following respects.

(i) Kalbfleisch writes, immediately after his formulation of his concept of an experimental ancillary, C_E, as an equivalence relation, 'This principle implies that in interpreting evidential meaning one *should* (italics added) condition . . .' Of course an equivalence relation does not literally imply a directive to make substitutions, still less a directive to make substitutions of one but not the other kind covered by the equivalence relation. If we adopt a concept represented by an equivalence relation $A = B$, then we may replace A by B wherever A occurs. But we must also allow B to be replaced by A wherever B occurs, if we are interested in exploring fully the theoretical implications and significance of the concept.

It was the adoption of an unqualified equivalence formulation of conditionality, and related concepts, which led, in my 1962 paper, to the monster of the likelihood axiom. What would be the precise theoretical form of a modified conditionality concept which might prove adequate with respect to theoretical content, and constructively compatible with typical uses of standard statistical methods? Kalbfleisch's 'should' is evidently based on an essential but tacit appeal to a desideratum of continuity with a tradition of statistical thought and practice, namely 'the operational approaches of both the Fisherian and the Neyman–Pearson schools', which he does not characterize fully in terms of his explicitly developed theoretical concepts.

(ii) Kalbfleisch's basic theoretical concept of an experimental ancillary seems to lose plausibility as a clear and satisfactory concept in the light of the following variation of his Example 1.

Suppose that the parameter μ of interest in that example is the unknown value of a physical constant in a system of otherwise completely known physical laws. It is typical of such systems that they have sufficiently general scope to characterize the behaviour of experimental set-ups of various different forms. Thus set-ups of different forms may be adopted as alternative experimental bases for measuring or estimating the unknown value, and the form of a set-up then has no intrinsic interest apart from its convenience for this purpose.

Suppose now that one such set-up is represented by the model E_1 in Kalbfleisch's Example 1. Suppose that an alternative set-up is represented by E_1', a two-stage experiment: Its first stage consists of taking a single observation U on a physical randomization device, the distribution of U having the completely known form given in Kalbfleisch's Example 1 for the random variable A. The second stage is based on a series of other physical set-ups, indexed by the possible values u of U. If $U = u$

is the result of the first stage, then set-up u is observed, resulting in an observation y_u of a random variable Y_u whose distribution is symmetrical about the unknown constant μ of interest, and has the same form as the distribution of $\frac{1}{2}(X_1+X_2)$, conditional on $X_1-X_2 = u$, where X_1 and X_2 are independent and each has the distribution (1) in Kalbfleisch's example. Writing U for Kalbfleisch's A, the random point (U, Y_U) has the same distributions, known except for the parameter μ, in E_1 and in E_1'. Any estimation method based on E_1 can be duplicated in its probability properties by a method based on E_1', and vice versa.

Hence there is evidently no basis for preference by an investigator as between the experimental set-ups represented, respectively, by E_1 and E_1'. Nor is there any evident basis for the investigator's using or interpreting the observed sample point (u, y) differently if it is obtained from set-up E_1' rather than from set-up E_1.

However, Kalbfleisch's basic concept classifies u as a mathematical ancillary if it was obtained from set-up E_1, but as an experimental ancillary if it was obtained from E_1'. It is disconcerting that this concept distinguishes cases which seem equivalent on all scientific and statistical grounds. Further, this suggests the question whether, in more complicated examples, Kalbfleisch's proposed rule based on this concept will sometimes give different results depending on which of two set-ups was used, even if those set-ups are scientifically and statistically equivalent in the sense illustrated by E_1 and E_1'.

REFERENCES

BASU, D. (1959). The family of ancillary statistics. *Sankhyā* **21**, 247–56.

BASU, D. (1964). Recovery of ancillary information. *Sankhyā* **26** A, 3–16.

BIRNBAUM, A. (1961). On the foundations of statistical inference. I. Binary experiments. *Ann. Math. Statist.* **32**, 414–35.

BIRNBAUM, A. (1962). On the foundations of statistical inference (with discussion). *J. Am. Statist. Assoc.* **57**, 269–326.

BIRNBAUM, A. (1970). On Durbin's modified principle of conditionality. *J. Am. Statist. Assoc.* **65**, 402–3.

BIRNBAUM, A. (1972). More on concepts of statistical evidence. *J. Am. Statist. Assoc.* **67**, 858–61.

COX, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* **29**, 357–72.

COX, D. R. (1971). The choice between alternative ancillary statistics. *J. R. Statist. Soc.* B **33**, 251–5.

COX, D. R. & HINKLEY, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.

DURBIN, J. (1970). On Birnbaum's theorem on the relation between sufficiency, conditionality, and likelihood. *J. Am. Statist. Assoc.* **65**, 395–8.

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

KALBFLEISCH, J. D. (1975). Sufficiency and conditionality. *Biometrika* **62**, 251–9.

LAKATOS, I. (1963–4). Proofs and refutations. *Brit. J. Phil. Sci.* **14**, 1–25, 120–39, 221–43, 296–342.

[*Received January* 1975]

Comments on paper by J. D. Kalbfleisch: Some personal comments on sufficiency and conditionality

By A. D. McLAREN

Department of Statistics, University of Glasgow

1. SUFFICIENCY AND CONDITIONALITY

Kalbfleisch (1975) rides to the rescue of those who believe in some form of sufficiency and conditionality but do not accept the likelihood principle. Birnbaum's (1962) result that his formulations S and C together imply the likelihood principle is a challenge to those statisticians. Kalbfleisch seeks to turn away the threat in two ways. He substitutes for S a weakened form, S', which is restricted to minimal experiments; for C he substitutes a strengthened form C_E which permits the mixing distribution to be unknown. The first of these moves seems to me faint-hearted and the second foolhardy.