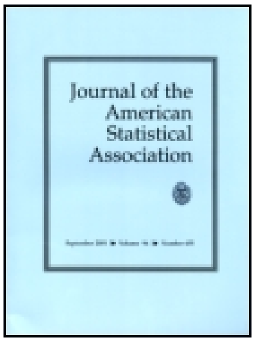

Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

On Birnbaum's Theorem on the Relation between Sufficiency, Conditionality and Likelihood

J. Durbin

To cite this article: J. Durbin (1970) On Birnbaum's Theorem on the Relation between Sufficiency, Conditionality and Likelihood, Journal of the American Statistical Association, 65:329, 395-398

To link to this article: https://doi.org/10.1080/01621459.1970.10481088

Published online: 05 Apr 2012.

© Journal of the American Statistical Association
March 1970, Volume 65, Number 329
Theory and Methods Section

On Birnbaum's Theorem on the Relation Between Sufficiency, Conditionality and Likelihood

J. DURBIN*

If the conditioning variable is required to depend only on the value of the minimal sufficient statistic, Birnbaum's proof fails.

Birnbaum's [1] demonstration that plausible principles of sufficiency and conditionality imply the likelihood principle has attracted considerable interest. However, although Birnbaum's sufficiency principle implies that, as a function of the observations, evidential meaning depends only on the minimal sufficient statistic, where this exists, the domain of application of his conditionality principle is not restricted to statistics which are minimally sufficient. It turns out that if the conditioning variable is required to be part of the minimal sufficient statistic, Birnbaum's proof fails. Consequently, if it is stipulated that any analysis or interpretation of the results should depend on the observations only through the value of the minimal sufficient statistic, Birnbaum's theorem is inapplicable.

Birnbaum's starting point is the concept of evidential meaning which is introduced by him in the following terms:

> "The present analysis of the first problem begins with the introduction of the symbol $Ev(E, x)$ to denote the *evidential meaning* of a specified instance (E, x) of statistical evidence; that is, $Ev(E, x)$ stands for the essential properties (which remain to be clarified) of the statistical evidence, as such, provided by the observed outcome x of the specified experiment E" [1, p. 270].

He goes on to define the concept of a mixture experiment as follows:

> "An experiment E is called a *mixture* (or a mixture experiment), with *components* $\{E_h\}$, if it is mathematically equivalent (under relabeling of sample points) to a two-stage experiment of the following form:
>
> (a) An observation h is taken on a random variable H having a fixed and known distribution G. (G does not depend on unknown parameter values.)
>
> (b) The corresponding component experiment E_h is carried out, yielding an outcome x_h" [1, p. 279].

The theory is based on the following two axioms denoted by Birnbaum as (S) and (C):

> "*Principle of Sufficiency* (S)*:* Let E be any experiment, with sample space $\{x\}$, and let $t(x)$ be any sufficient statistic (not necessarily real-valued). Let E' denote the derived experiment, having the same parameter space, such that when any outcome x of E is observed the corresponding outcome $t=t(x)$ of E' is observed. Then for each x, $Ev(E, x)=Ev(E', t)$, where $t=t(x)$" [1, p. 278].
>
> "*Principle of conditionality* (C)*:* If an experiment E is (mathematically equivalent to) a mixture G of components $\{E_h\}$, with possible outcomes (E_h, x_h), then

* J. Durbin is professor of statistics, London School of Economics, University of London. His published work is mainly in the fields of regression, time series and sample surveys. The author wishes to thank P. Armitage, A. Birnbaum, V. P. Godambe, L. J. Savage, H. Scheffé, and A. Stuart for helpful comments on an earlier draft of this note.

$$Ev(E,(E_h, x_h)) = Ev(E_h, x_h).$$

That is, the evidential meaning of any outcome (E_h, x_h) of any experiment E having a mixture structure is the same as: the evidential meaning of the corresponding outcome x_h of the corresponding component experiment E_h, ignoring otherwise the over-all structure of the original experiment E" [1, p. 279].

Birnbaum's theorem, derived as Lemma 2 [1, p. 284], states that (*S*) and (*C*) imply, and are implied by, the likelihood principle, which states that the evidential meaning of an observed outcome x is characterised completely by the likelihood function calculated from x [1, p. 283].

The first point to note is that (*S*) implies that $Ev(E, x)$ depends only on the minimal sufficient statistic, where this exists. To show this, let $t(x)$ denote the sufficient statistic referred to in (*S*) and let $t'(x)$ denote a minimal sufficient statistic. Let E'' denote the derived experiment, having the same parameter space as E', such that when the outcome $t(x)$ of E' is observed the corresponding outcome $t'(x)$ of E'' is observed. By (*S*), $Ev(E, x) = Ev(E', t(x)) = Ev(E'', t'(x))$. It follows that, as a function of the outcome x, $Ev(E, x)$ depends only on the minimal sufficient statistic $t'(x)$. (We may speak of *the* minimal sufficient statistic without loss since any minimal sufficient statistic is a function of every other minimal sufficient statistic.) This implies that if $t'(x)$ is preserved the remaining record of the results of the experiment can be destroyed without in any way affecting the determination of $Ev(E, x)$.

Since evidential meaning depends only on the minimal sufficient statistic it would seem reasonable to require that any analysis or interpretation of the results of the experiment should depend only on the value of the minimal sufficient statistic. This leads naturally to the requirement that the domain of applicability of (*C*) should be restricted to components of the minimal sufficient statistic. We are therefore led to the following modified form of (*C*).

Modified principle of conditionality (C'): If an experiment E is (mathematically equivalent to) a mixture G of components E_h, with possible outcomes (E_h, x_h), where h depends only on the value of the minimal sufficient statistic, then

$$Ev(E, (E_h, x_h)) = Ev(E_h, x_h).$$

Within the framework of Birnbaum's terminology this seems to be consistent with what most statisticians mean by conditional inference. We now show that Birnbaum's proof fails when (*C*) is replaced by (C').

Let E be a mixture experiment as defined by Birnbaum and consider two outcomes, $h=i$, $x=x_i$ and $h=j$, $x=x_j$ where the conditional densities of x_i, x_j given i and j are $f_i(x, \theta)$ and $f_j(x, \theta)$. Suppose that x_i and x_j lead to the same likelihood function, i.e., $f_i(x_i, \theta) = cf_j(x_j, \theta)$ for all θ where c may depend on x_i and x_j but not on θ. Denote the density of the known distribution G of h by g_h. The conditional probability that $h=i$ and $x=x_i$ given that $h=i$, $x=x_i$ or $h=j$, $x=x_j$ is

$$\frac{g_i f_i(x_i, \theta)}{g_i f_i(x_i, \theta) + g_j f_j(x_j, \theta)} = \frac{g_i c}{g_i c + g_j},$$

which does not depend on θ. Consequently, given that $h=i$, $x=x_i$ or $h=j$, $x=x_j$, the actual value of h (i or j) cannot be part of the minimal sufficient statistic. Thus the statement

$$Ev(E, (E_h, x_h)) = Ev(E_h, x_h), \tag{1}$$

which follows from (C), does not follow from (C').

The proof of Birnbaum's theorem depends crucially on the use of (1) in a situation of the kind we have been considering. Birnbaum considers two experiments E and E' with outcomes x, y having densities $f(x, \theta)$, $g(y, \theta)$ on sample spaces S, S'. He takes a particular pair of outcomes x', y', which lead to the same likelihood function, i.e., $f(x', \theta) = cg(y', \theta)$ for all θ where c is a positive constant and introduces a (hypothetical) mixture experiment E^* whose components are just E and E', taken with equal probabilities. Birnbaum then invokes (C) to justify the assertions

$$\begin{aligned} Ev(E^*, (E, x)) &= Ev(E, x) \text{ for each } x \in S, \text{ and} \\ Ev(E^*, (E', y)) &= Ev(E', y) \text{ for each } y \in S' \end{aligned} \tag{2}$$

((5.1) in [1]). On applying (2) to the cases $x=x'$, $y=y'$ and using (S) his theorem follows.

In order to examine the situation from the standpoint of (C'), let a be an indicator variable such that $a=0$ if E is taken and $a=1$ if E' is taken. The outcome of E^* can then be written as $(0, x)$ or $(1, y)$ according as E or E' is taken. The conditional probability of $(0, x')$ given $\{(0, x') \text{ or } (1, y')\} = 1/2f(x', \theta) / \{1/2f(x', \theta) + 1/2g(y', \theta)\} = c/(1+c)$ which is independent of θ. Thus for the particular outcomes x' and y', the value of a cannot be part of the minimal sufficient statistic. Under (C') there is therefore no justification for the assertions (2) for the outcomes $x=x'$ or $y=y'$. Since these assertions are essential for Birnbaum's proof his theorem fails when (C) is replaced by (C').

It must not be thought that the above discussion can be dismissed as mere technical quibbling. On the contrary, I believe that the substitution of (C') for (C) involves a serious questioning of Birnbaum's approach to conditionality and of the way he relates it to sufficiency. It should be noted that his theorem cannot be "disproved" in the ordinary sense owing to the fact that $Ev(E, x)$ is not defined. If $Ev(E, x)$ is interpreted as likelihood from the outset then the theorem is trivially true. It is for this reason that the approach of this note has been to indicate the point at which his argument breaks down under (C') rather than to attempt to present counterexamples.

I wish to emphasize that my main purpose has not been to argue the case for the adoption of (C') as a basic principle but rather, granted its plausibility, to demonstrate the consequences of (C') for Birnbaum's theory of conditionality.

Some of the points made in this note are criticized by Savage [4] in the note immediately following it. Other aspects of Birnbaum's theorem are discussed by Fraser [2] and Hájek [3].

REFERENCES

[1] Birnbaum, A., "On the Foundations of Statistical Inference," *Journal of the American Statistical Association*, 57 (1962), 269–306.

[2] Fraser, D. A. S., "On the Sufficiency and Likelihood Principles," *Journal of the American Statistical Association*, 58 (1963), 641–7.
[3] Hájek, J., "On Basic Concepts of Statistics," *Proceedings of the Fifth Berkeley Symposium, Vol. I* (1967), 139–62.
[4] Savage, Leonard J., "Comments on a Weakened Principle of Conditionality," *Journal of the American Statistical Association*, 65 (1970).