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

Comments on a Weakened Principle of Conditionality

LEONARD J. SAVAGE*

Continuity suggests that the conditioning variable not be required to depend only on the values of the minimal sufficient statistic.

In kindness to me and for the convenience of readers, J. Durbin has urged that his critical note [2] on the principle of conditionality be followed immediately by this dissent.

The idea that the principle of conditionality (C) ought to be weakened to (C') because all the data of an experiment beyond the minimal sufficient statistic are superfluous seems to me like the plausible and frustrating idea of the geometry pupil who, since lines of construction cannot affect the truth of a theorem, concludes that they ought not be referred to in its demonstration. Each may judge for himself whether this simile is well taken; I find it hard to expand upon fruitfully and shall not pursue it further here. Rather, most of my space will be devoted to showing that certain continuity assumptions that seem compelling restore the strength of (C) to (C'). As Durbin implies, the subject does not seem to admit of an objectively final argument.

Briefly, even if (C) and (C') lead to different conclusions about an experiment E, then E can be replaced, for any practical purpose, by a variant $\hat{E}$ that differs from E only microscopically and is such that (C) and (C') as applied to $\hat{E}$ are absolutely identical and lead to conclusions about $\hat{E}$ that differ only microscopically from the conclusions of (C) applied to E. For some, that line of thought will seem transparent, and they may find the following extended illustration tedious. But some such spelling out seems necessary to make the argument precise and convincing, especially to those who are inclined to see an escape from the likelihood principle in the distinction between (C) and (C'). An example seems more valuable here than a piece of general theory. Those of theoretical temperament will see how the example suggests general theory; others might see nothing at all in general theory unillustrated.

Consider a particular pair of experiments E and E' such as that referred to in the paragraph following (1) of Durbin's note. Specifically, let the experiment E consist in observing the number x of Bernoulli trials, each of probability θ, required to attain 100 successes, and let E' consist in observing the number y of successes in 1,000 such Bernoulli trials with the same θ. The corresponding "densities" are

$$f(x, \theta) = \binom{x-1}{99}\theta^{100}(1-\theta)^{x-100}$$

(for $x=100, 101, \cdots$) and

* Leonard J. Savage is Eugene Higgins professor and chairman, Department of Statistics, Yale University. This research was supported by the Army, Navy, Air Force and NASA under a contract administered by the Office of Naval Research.

$$g(y, \theta) = \binom{1{,}000}{y} \theta^{y}(1 - \theta)^{1{,}000-y}$$

(for $y=0, 1, \cdots, 1{,}000$).

For $x'=1{,}000$, $y'=100$, and all θ, $f(x', \theta)=(100/1{,}000)g(y', \theta)$. Therefore, Birnbaum's argument based on (C) would imply that (E, 1,000) and (E', 100) have the same evidential meaning. But, as Durbin emphasizes, the weaker principle (C'), even in the presence of the sufficiency principle (S), is not adequate to support that (or presumably any other) argument leading to this conclusion of equivalence. In the presence of certain continuity assumptions, (C') with (S) does imply (C). Thus, to one who finds these continuity assumptions compelling, rejecting (C) in favor of (C')—whether that is justified or not—is without any net effect.

To illustrate, consider a third experiment $\hat{E}'$, practically indistinguishable from E', though mathematically distinct from it. Formally, $\hat{E}'$ has a density $\hat{g}(y, \theta)$ such that, for all θ properly between 0 and 1 and for all y, the likelihood ratio $g(y, \theta)/\hat{g}(y, \theta)$ is between $1-\epsilon$ and $1+\epsilon$ for any small ϵ of your choice, though the likelihood ratio is not constant for $y=100$. Such experiments are easy to construct; one colorful way, which may make the ensuing argument seem more familiar, is to sample without replacement from a suitable finite population that is enormous (compared to $1/\theta$ and $1/(1-\theta)$).

In any given context, are there not values of ϵ so small that no statistician would think it important to react differently to (E', y) and ($\hat{E}'$, y)? The two data thus constitute practically the same evidence. This may be suggestively recorded thus:

$$Ev(E', y) \simeq Ev(\hat{E}', y). \tag{1}$$

How common and right it is to say that sampling with, and sampling without, replacement are practically equivalent when the finite population is large compared to the sample size! Solid though (1) seems to me, it represents an assumption and must be judged by the reader himself, as must a somewhat less familiar continuity assumption to be introduced a little later.

Incidentally some seem to find the notion of a function, Ev, the value of which is the evidential meaning of the outcome x to the experiment E so disturbing as to impede further discussion. They, like me, might prefer to begin with the notion that sometimes the outcome x of experiment E cannot justifiably provoke any different statistical reaction (such as a decision or an inference concerning θ) from those provoked by x' and E'. Such a pair might be called evidentially equivalent, though just which pairs are equivalent is still under discussion. It is a logical truism that an equivalence relation can be described by the function that associates each object with its equivalence class, so that (E, x) *ev eq* (E', x') and $Ev(E, x)=Ev(E', x')$ are but two ways of saying the same thing; your immediate family is the class of people to whom you are closely related, and your graduating class is the class of people indistinguishable from you in year and school of graduation.

Consider now the mixture experiment $\hat{E}^*$, in which E or $\hat{E}'$ is executed, each with probability 1/2. In contrast with what Durbin pointed out about E^*, the element of the mixture that happens to be executed and its outcome do constitute a minimal sufficient statistic for $\hat{E}^*$. Therefore, (C') now is applicable to conclude that

$$Ev(\hat{E}^*, (E, 1{,}000)) = Ev(E, 1{,}000), \tag{2}$$

$$Ev(\hat{E}^*, (\hat{E}', 100)) = Ev(\hat{E}', 100). \tag{3}$$

In $\hat{E}^*$, the two possible data $(E, 1{,}000)$ and $(\hat{E}', 100)$ have almost, but not quite, proportional probabilities as a function of θ, that is, they have almost the same likelihood. Were the proportionality exact, the principle of sufficiency (S), which no one known to me now rejects, would imply that these two data have the same evidential meaning. Since the approximate proportionality becomes arbitrarily close with decreasing ϵ, the two evidential meanings must, I submit, be nearly the same. For example, recommending such macroscopically different point estimates as 99/999 and 100/1,000 (suggested by the unbiased-estimate tradition) for the two potential data in one experiment does not seem appropriate.

If it is granted that

$$Ev(\hat{E}^*, (E, 1{,}000)) \simeq Ev(\hat{E}^*, (\hat{E}', 100)), \tag{4}$$

then

$$Ev(E, 1{,}000) \simeq Ev(E', 100), \tag{5}$$

as follows directly from (2), (4), (3), and (1) in that order. But since E and E' do **not** change with ϵ, the arbitrarily close approximation asserted by (5) must really be exact equivalence, which is the conclusion to be demonstrated.

Perhaps the continuity assumptions will not be compelling to all readers; but, mathematically, each can now see, I hope, according to his own standards of generality and precision how (C') and (S) do imply (C) in the presence of such assumptions.

Barnard, Jenkins, and Winsten [1] in deriving the likelihood principle from mixtures use the idea that nearly identical experiments should provoke nearly identical reactions, but their application of the idea seems quite different from that in this note.

REFERENCES

[1] Barnard, G. A., Jenkins, G. M., and Winsten, C. B., Series (with Discussion)," *Journal of the Royal Statistical Society, Series A*, 125 (1962), 321–72.

[2] Durbin, J., "On Birnbaum's Theorem on the Relation between Sufficiency, Conditionality and Likelihood," *Journal of the American Statistical Association*, 65 (March 1970), 395–8.