

# Discussion of "On the Birnbaum Argument for the Strong Likelihood Principle" by D.G. Mayo

**Jan F. Bjørnstad**

The goal of this paper is to provide a new clarification and critique of Birnbaum's argument for showing that principles of sufficiency and conditionality entail the (strong) likelihood principle (LP). I must admit I do not find that the paper provides such a new clarification of the criticism of Birnbaum's argument. Rather, much of the criticism in the paper goes back to arguments made in the 70's and 80's by several authors, e.g., Durbin (1970), Kalbfleisch (1975), Cox (1978) and Evans et al (1986). This critique has been discussed by several statisticians with an opposing view, see Berger & Wolpert (1988) and Bjørnstad (1991).

I will concentrate my discussion on what seems to be the most important contention in the paper, that the sufficient statistic in Birnbaum's proof erases the information as to which experiment the data came from and hence that the weak conditionality principle (WCP) cannot be applied, see, e.g., sections 5.2 and 7.

As I understand it, this is a misunderstanding of the proof. For one thing, it seems that only the observation  $x_2$  from the experiment  $E_2$  is considered in the mixture experiment instead of the correct  $(E_2, x_2)$ . The observations in a mixture experiment is always of the form  $(E_h, x_h)$ - *never* as only  $x_h$ .

Other arguments leading up to this contention seem to rest on a misunderstanding of the sufficiency considerations in the proof, that given an observation from a certain experiment the result in an unperformed experiment is to be reported, see e.g. sections 2.4, 5.1. I find that this is definitely not the case. To be specific, the author considers the following proof in the discrete case:

Let  $(j, x_j)$  indicate that Experiment  $E_j$  was performed with data  $x_j, j=1, 2$ . Assume the data values  $x_1^0$  and  $x_2^0$  have proportional likelihoods from experiments  $E_1$  and  $E_2$  respectively. Then the sufficient statistic in the mixture experiment used in Birnbaum's proof is given by:

$$T(j, x_j) = (j, x_j) \text{ if } (j, x_j) \neq (1, x_1^0), (2, x_2^0)$$

$$T(1, x_1^0) = T(2, x_2^0) = (1, x_1^0).$$

Mayo claims that  $T(1, x_1^0) = T(2, x_2^0) (= c)$  implies that the weak conditionality principle (WCP) is violated. Now, one should note that the proof works with any  $c \neq (1, x_1^0)$  and  $c \neq (j, x_j)$   $j = 1, 2$  and all  $x_j$ . When  $E_2$  is performed and  $x_2^0$  is the result then the evidence should only depend on  $E_2$  and  $x_2^0$ , and not on a result of an unperformed experiment  $E_1$ . This is, of course, correct, but it does not depend on a result in  $E_1$ . (Actually  $E_1$  is not an unperformed experiment either. We comment on this issue later.) By letting  $c = (1, x_1^0)$  it seems so, but we see by choosing a  $c \neq (1, x_1^0)$  it is not the case. So WCP is not violated. The sufficient statistic simply takes the same value for these two results of the mixture experiment. It has nothing to do with WCP.

So Birnbaum's proof does not require that the evidential support of a known result should depend on the result of an unperformed experiment. It follows that the main contention in the paper seems to rest on a lack of understanding of the basics of the proof of Birnbaum's theorem. In fact, it is possible to do the proof even more general. One can show, for a given experiment, see, e.g., Cox and Hinkley (1974) and Bjørnstad (1996), that if two likelihoods are proportional for two possible observations in the same experiment, there exists a minimal sufficient statistic with the same value for the two observations. This holds both for discrete and continuous models.

To make the sufficiency argument clearer, consider a mixture of a binomial experiment  $E_1$  and a negative binomial experiment  $E_2$  where the observations are  $x_1 =$  number of successes in 12 trials and  $x_2 =$  number of trials until 3 successes. If  $x_1^0 = 3$  and  $x_2^0 = 12$ , then the likelihoods are proportional. A natural choice of the sufficient statistic  $T$  in the mixture experiment in Birnbaum's proof has

$$T(1, x_1^0) = T(2, x_2^0) = 3/12, \text{ the proportion of successes in either case.}$$

Clearly then, the value of  $T$  from experiment  $E_2$  does not depend on the result from  $E_1$ .

As already mentioned, the author claims that the sufficient statistic  $T$  in the proof of Birnbaum's result has the effect of erasing the index of the experiment. Moreover it is claimed that inference based on  $T$  is to be computed over the performed and unperformed experiments  $E_1$  and  $E_2$ . As we have shown, this is simply not the case. It should also be mentioned that statistically the proof simply considers two instances of performing the mixture experiments resulting in proportional likelihoods and really has nothing to do with considering unperformed experiments.

Let me also mention that the author's premise in section 5 is not correct. The starting point is *not* that we have an arbitrary outcome of one single experiment, but rather that two experiments have been

performed about the same parameter resulting in proportional likelihoods. So Birnbaum does not enlarge a known single experiment but construct a mixture of the two performed experiments. There is really *no unperformed* experiment here. In a sense, one may regard the paper by Mayo as actually not discussing the LP at all.

It should be clear that I find that the main contention in the paper does not hold when maintaining the original meaning of the principles of sufficiency, conditionality and likelihood. Other comments made in this paper referring to various authors in the 70's and 80's are a different matter. However, I do not find any *new* clarification of Birnbaum's fundamental theorem in this paper. For example regarding sufficiency, it is necessary to restrict the application of sufficiency to non-mixture experiments, as Kalbfleisch (1975) did, in order to invalidate Birnbaum's result. Berger and Wolpert (1988) argues, I think, convincingly against such a restriction. See also Bjørnstad (1991).

Let me end this discussion by making clear the following fact: It is obviously clear that frequentistic measures may, and typically do, violate LP. This is true as far as it comes to analysis of the actual data we observe. But a major point here is that the LP does not say that one should not be concerned with how the methods do when used repeatedly. LP is simply *not about method evaluation*. Evaluation of methods is still important. So LP says in essence that frequentistic considerations are not *sufficient* for evaluating the uncertainty and reliability in the statistical analysis of the actual data, see also Bjørnstad (1996) for a discussion on this issue.

## REFERENCES

- BERGER, J.O. and WOLPERT, R.L. (1988). *The Likelihood Principle*, (2<sup>nd</sup> ed.) Vol.6. Lecture Notes-Monograph Series. Hayward, California: Institute of Mathematical Statistics.
- BJØRNSTAD, J.F. (1991). Introduction to Birnbaum (1962): On the foundations of statistical inference. In *Breakthroughs in Statistics*, vol.1, 461- 477. Ed. Kotz, S. and Johnson, J. Springer Series in Statistics, New York: Springer-Verlag.
- BJØRNSTAD, J.F. (1996). On the generalization of the likelihood Function and the likelihood principle. *J. Amer. Stat. Assoc.* **91** 791 – 806.
- COX, D.R. (1978). Foundations of statistical inference: The case for eclectism. *Australian J. Statist.* **20** 43-59.

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

DURBIN, J. (1970). On Birnbaum's theorem in the relation between sufficiency, conditionality and likelihood. *J. Amer. Stat. Assoc.* **65** 395 – 398.

EVANS, M., FRASER, D.A.S., and MONETTE, G. (1986). On principles and arguments to likelihood (with discussion). *Canadian J. Statist.* **14** 181 – 199.

KALBFLEISCH, J.D. (1975). Sufficiency and conditionality (with discussion). *Biometrika* **62** 251 – 268.