In Defense of the Likelihood Principle

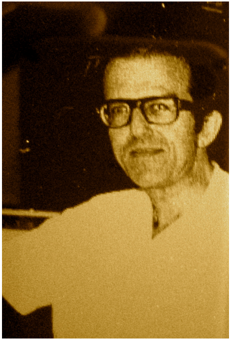
James O. Berger, Víctor Peña

Department of Statistical Science
Duke University

ISBA 2016

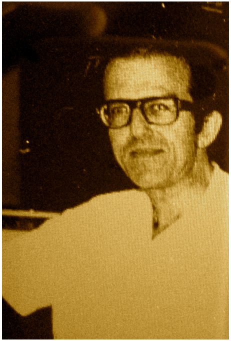
Birnbaum's Theorem

- Birnbaum (1962) proves that the Likelihood Principle (LP) is implied by the weak sufficiency principle (WSP) and weak conditionality principle (WCP).
- **Huge result!** WSP and WCP would seem appealing to most statisticians, but LP has strong implications.



Today's Talk

- LP doesn't allow *p*-values, reference priors, dependence on stopping rules, etc.
- There is a vast literature that discusses the theorem and its implications (cf. Berger and Wolpert (1988))
- Recently, Evans (2013) and Mayo (2014) have received considerable attention.
- **Our goal is reviewing and discussing their content.**



Statistical Principles

WCP, AP, WSP

- **Weak Conditionality Principle** (WCP): If you choose your experiment by flipping a coin, your inferences should only depend on the observed experiment.
- **Ancillarity Principle** (AP): Your inferences shouldn't change after observing (conditioning on) the value of a statistic whose distribution doesn't depend on unknown parameters.
- **Weak Sufficiency Principle** (WSP): If T is a sufficient statistic and x_1 and x_2 are outcomes such that $T(x_1) = T(x_2)$, your inferences upon observing x_1 or x_2 should be the same.

LP and Birnbaum's Theorem

- **Likelihood Principle** (LP): If two experiments for the same parameter have proportional likelihoods (given the observed data), the inferences to be made from the experiments should be the same.

Birnbaum's theorem

WCP and WSP imply LP.

Evans (2013): Review and Discussion

Electronic Journal of Statistics
Vol. 7 (2013) 2645–2655
ISSN: 1935-7524
DOI: 10.1214/13-EJS857

What does the proof of Birnbaum's theorem prove?

Michael Evans

Department of Statistics
University of Toronto
e-mail: mevans@utstat.utoronto.ca
url: www.utstat.utoronto.ca/mikevans

Abstract: Birnbaum's theorem, that the sufficiency and conditionality principles entail the likelihood principle, has engendered a great deal of controversy and discussion since the publication of the result in 1962. In particular, many have raised doubts as to the validity of this result. Typically these doubts are concerned with the validity of the principles of sufficiency and conditionality as expressed by Birnbaum. Technically it would seem, however, that the proof itself is sound. In this paper we use set theory to formalize the context in which the result is proved and show that in fact Birnbaum's theorem is incorrectly stated as a key hypothesis is left out of the statement. When this hypothesis is added, we see that sufficiency is irrelevant, and that the result is dependent on a well-known flaw in conditionality that renders the result almost vacuous.

AMS 2000 subject classifications: Primary 62A01; secondary 62F99.
Keywords and phrases: Sufficiency, conditionality, likelihood, relations, equivalence relations.

Received February 2013.

Evans (2013): Main Result

- Evans (2013) considers AP instead of WCP, but his main objection carries over if we consider WCP.
- Define WSP, WCP, and LP as set relations S, C, and L, respectively (for example, $(E_1, x_1) \sim_S (E_2, x_2)$ whenever they are "equivalent" according to WSP).
- Then, $S \cup C \neq L$.

Interpreting the Main Result

What does $S \cup C = L$ mean?

If $S \cup C = L$, the following statement would be true. Let E_1 and E_2 be experiments about the same θ, and assume that x_1^* and x_2^* are such that $f_\theta^1(x_1^*) \propto f_\theta^2(x_2^*)$. Then, $\mathrm{Ev}(E_1, x_1^*) = \mathrm{Ev}(E_2, x_2^*)$ **by an application of either WSP or WCP alone**.

- Clearly, $S \cup C = L$ is **not** implied by Birnbaum's proof (and it's actually false!).

$$\boxed{\mathrm{Ev}(E_1, x_1^*) \overset{\text{WCP}}{=} \mathrm{Ev}(E, (1, x_1^*)) \overset{\text{WSP}}{=} \mathrm{Ev}(E, (2, x_2^*)) \overset{\text{WCP}}{=} \mathrm{Ev}(E_2, x_2^*)}$$

Conclusion in Evans (2013)

Evans (2013)

We have shown that the proof in [Birnbaum (1962)] did not prove that S and C lead to L. [...] The statement of Birnbaum's theorem in prose should have been: if we accept the relation S and we accept the relation C and we accept all the equivalences generated by S and C together, then this is equivalent to accepting L. The essential flaw in Birnbaum's theorem lies in excluding this last hypothesis from the statement of the theorem.

- What are "all the equivalences generated by S and C together"?

Equivalences and "Transitivity"

- The equivalences generated by S and C are simply "chains" of applications of S and C.
- Is this assumption *really* left out of the proof? To us, the notation "Ev" and the "$=$" in the definitions of the principles imply the "key assumption".
- And if it seems ambiguous, the proof itself makes it clear.

Is transitivity a strong condition?

- Evans (2013) argues against the extension of AP to an equivalence relation because it is “essentially equivalent to saying that it doesn’t matter which maximal ancillary we condition on and it is unlikely that this is acceptable to most frequentist statisticians”.
- If you truly believe in your principles, why shouldn’t you be able to apply them in any order you want, and as many times as you want?
- We wonder what Evans thinks of the proof of “WCP + WSP implies LP.” Is combining applications of WCP and WSP reasonable?

Mayo (2014): Review and Discussion

Statistical Science
2014, Vol. 29, No. 2, 227–239
DOI: 10.1214/13-STS457

On the Birnbaum Argument for the Strong Likelihood Principle[1]

Deborah G. Mayo

Abstract. An essential component of inference based on familiar frequentist notions, such as p-values, significance and confidence levels, is the relevant sampling distribution. This feature results in violations of a principle known as the strong likelihood principle (SLP), the focus of this paper. In particular, if outcomes $\mathbf{x}^*$ and $\mathbf{y}^*$ from experiments E_1 and E_2 (both with unknown parameter θ) have different probability models $f_1(\cdot)$, $f_2(\cdot)$, then even though $f_1(\mathbf{x}^*;\theta) = cf_2(\mathbf{y}^*;\theta)$ for all θ, outcomes $\mathbf{x}^*$ and $\mathbf{y}^*$ may have different implications for an inference about θ. Although such violations stem from considering outcomes other than the one observed, we argue this does not require us to consider experiments other than the one performed to produce the data. David Cox [*Ann. Math. Statist.* **29** (1958) 357–372] proposes the Weak Conditionality Principle (WCP) to justify restricting the space of relevant repetitions. The WCP says that once it is known which E_i produced the measurement, the assessment should be in terms of the properties of E_i. The surprising upshot of Allan Birnbaum's [*J. Amer. Statist. Assoc.* **57** (1962) 269–306] argument is that the SLP appears to follow from applying the WCP in the case of mixtures, and so uncontroversial a principle as sufficiency (SP). But this would preclude the use of sampling distributions. The goal of this

Basic Notation: Methods and Inference

- Mayo (2014) distinguishes between “methods” and “informative inference”.

New Notation: Methods and “Inferences”

- $\mathcal{M}(E, x)$: Result of a applying a a statistical method when E and x are taken as inputs.
- $\mathcal{I}(E', x')$: Informative inference (conclusion/decision, etc.) that is made after seeing data $X' = x'$ from E'.

- **NB:** $\mathcal{I}(E', x')$ may not be equal to $\mathcal{M}(E, x)$.

Principles in Mayo (2014)

WCP and WSP as in Mayo (2014), with our notation

- **WCP2:** Given $(E_{mix}, (j, x_j))$, $\mathcal{I}(E_{mix}, (j, x_j)) = \mathcal{M}(E_j, x_j)$.
- **WSP2:** If there exists a sufficient statistic T for θ and $T(x) = T(x')$, then $\mathcal{M}(E, x) = \mathcal{M}(E, x')$.

- With these definitions, we can construct examples where WCP2 and WSP2 are respected, but LP isn't.

Counterexample: Binomial vs Negative Binomial

Example in Mayo (2010)

- Let E_1 and E_2 be Bin(n, θ) and NB(r, θ) experiments for the same θ.
- Let E_{mix} denote the "coin-flip mixture experiment" between E_1 and E_2.
- Suppose **both** E_1 and E_2 yield $x = (n - r, r)$, where $n - r$ and r are the number of successes and failures observed after performing E_j.

Example in Mayo (2010) (cont.)

- $\mathcal{M}(E, x)$ is the (usual) one-sided p-value:

$$\begin{aligned}\mathcal{M}(E_1, x) &= P(\text{Bin}(n, \theta_0) \geq n - r)\\ \mathcal{M}(E_2, x) &= P(\text{NB}(r, \theta_0) \geq n - r)\\ \mathcal{M}(E_{\text{mix}}, x) &= \frac{1}{2}[\,P(\text{Bin}(n, \theta_0) \geq n - r) + \,P(\text{NB}(r, \theta_0) \geq n - r)],\end{aligned}$$

- Inference is made using the rule $\mathcal{I}(E_j, x) = \mathcal{M}(E_j, x)$ and $\mathcal{I}(E_{\text{mix}}, (j, x)) = \mathcal{M}(E_j, x)$.
- For any **given** experiment, the number of successes is suff. for θ and the p-value doesn't change for any value of x with the same suff. stat.
- It follows that WCP2 and WSP2 do not imply LP.

Example: Some Comments (I)

- WSP2 only applies to the method (given some inputs), not the the final inferences.
- As a result, reporting the conditional *p*-value in the example above isn't a violation of WSP2.
- If WSP2 is defined solely in terms of $\mathcal{I}$, WSP2 and WCP2 imply LP (in the Example we discussed, the existence of a suff. statistic requires that the inferences from E_{mix} and E_j be equal).
- The same happens if WCP2 were defined in terms of $\mathcal{M}$ (*p*-values as a “method” would be precluded).

Example: Some Comments (II)

- The notation "Ev(E, x)" is supposed to denote our "inference after seeing E from x", so the definitions should probably be written in terms of $\mathcal{I}$ alone.
- In any case, that distinction is not made in Birnbaum (1962).

Conclusions

- Evans (2013) claims that a notion of “transitivity of inferences” is a strong condition, but it seems that his criticisms apply to AP as a principle (and not “transitivity”).
- Mayo (2010) makes a distinction between “methods” and “informative inferences”. This distinction is not made in Birnbaum (1962).
- Conditional inference from a “*p*-value” perspective is hard.

References I

Berger, J. O. and Wolpert, R. L. (1988). The likelihood principle. *Lecture notes-Monograph series*.

Birnbaum, A. (1962). On the foundations of statistical inference. *Journal of the American Statistical Association*, 57(298):269–306.

Evans, M. (2013). What does the proof of Birnbaum's theorem prove? *Electronic Journal of Statistics*, 7:2645–2655.

Mayo, D. G. (2010). An error in the argument from conditionality and sufficiency to the likelihood principle. *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*, page 305.

Mayo, D. G. (2014). On the Birnbaum argument for the strong likelihood principle. *Statistical Science*, 29(2):227–239.