Birnbaum’s confidence principle

Ofer Vitells

Statistics miniworkshop at CERN , February 2013

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 298 JUNE, 1962 *Volume 57*

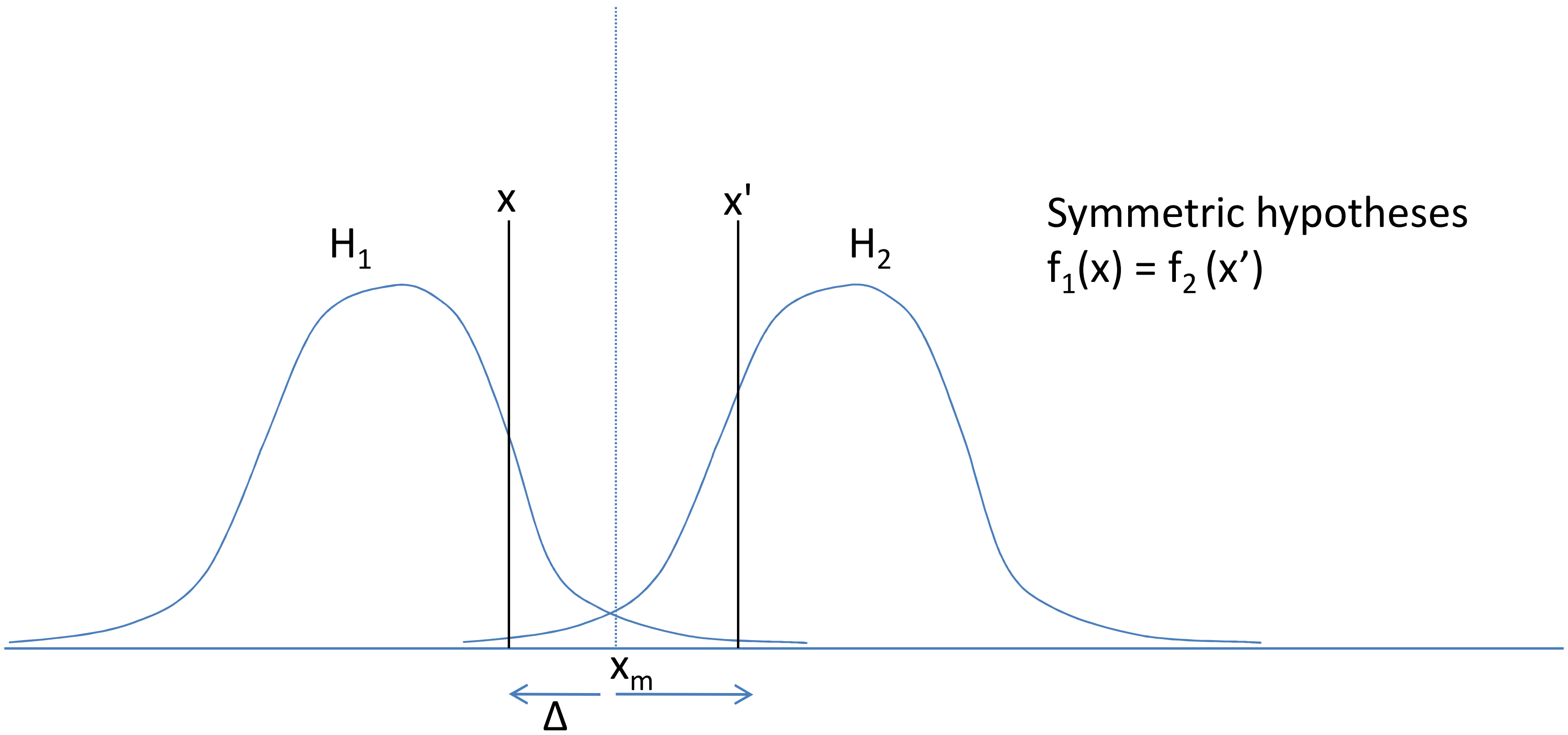
ON THE FOUNDATIONS OF STATISTICAL INFERENCE

ALLAN BIRNBAUM
New York University

- Birnbaum is most well known for his proof that conditionality (conditioning on ancillary statistics) implies the Likelihood principle

L. J. SAVAGE: Without any intent to speak with exaggeration or rhetorically, it seems to me that this is really a historic occasion. This paper is a landmark in statistics because it seems to me improbable that many people will be able to read this paper or to have heard it tonight without coming away with considerable respect for the likelihood principle.

Simple example of conditionality ==> LP



Symmetric hypotheses
$f_1(x) = f_2\,(x')$

$\Delta = |x - x_m|$ is an ancillary statistic

$P_{H1}(\, x \mid \Delta\,) = f1(x) / (f1(x) + f1(x')) = 1/(1 + f2(x)/f1(x)) = 1/(1 + LR(x))$

The Confidence Principle

- It would also be useful to define the basic principle that guides the use and interpretation of frequentist quantities in statistical inference (i.e. how they are related to evidence)

- Birnbaum suggested the following general principle:

(Conf): A concept of statistical evidence is not plausible unless it finds 'strong evidence for H_2 as against H_1' with small probability (α) when H_1 is true, and with much larger probability ($1-\beta$) when H_2 is true.

Birnbaum, 1977

(Conf): A concept of statistical evidence is not plausible unless it finds 'strong evidence for H_2 as against H_1' with small probability (α) when H_1 is true, and with much larger probability ($1-\beta$) when H_2 is true.

- Note that this is formulated as a **necessary** but **not** necessarily sufficient condition,
  i.e. small $\alpha/(1-\beta)$ does not imply strong evidence,
  but strong evidence implies small $\alpha/(1-\beta)$.

- Does not imply that additional requirements (conditionality, etc.) are unnecessary.

(Conf): A concept of statistical evidence is not plausible unless it finds 'strong evidence for H_2 as against H_1' with small probability (α) when H_1 is true, and with much larger probability ($1-\beta$) when H_2 is true.

- Classical Neyman-Pearson testing might equivalently be said to be based on the principle “strong evidence ==> small α”
  (but not “strong evidence <==> small α”)
- This is not in contradiction to Birnbaum’s C.P.

Motivation

Suppose an experiment has the following structure :

Simple hypotheses:	H_1,	H_2
Possible decisions:	d_1 ="reject H_1"	d_2 ="accept H_1"
Error probabilities:	$\alpha = \text{Prob}\,[d_1 \mid H_1]$,	$\beta = \text{Prob}\,[d_2 \mid H_2]$

- The overall procedure is equivalent to a Bernoulli trial with two possible outcomes (d_1, d_2) .
- The likelihood ratio of the outcome "reject H1" is $P(d_1 \mid H_1)/P(d_1 \mid H_2) = \alpha/(1-\beta)$.
- The likelihood principle therefore implies that inference (i.e. strength of evidence associated to "reject H_1") be a function of $\alpha/(1-\beta)$.

Motivation

- In a simple-vs-simple hypotheses case, the following inequality holds :

$$(1) \qquad \alpha \leqslant \frac{\alpha}{1-\beta} \leqslant \frac{d\alpha}{d\beta} = \text{likelihood ratio} \leqslant \frac{1-\alpha}{\beta} \leqslant \frac{1}{\beta}.$$

- In this case the LP and CP can be satisfied at the same time, e.g.: reject H_1 if LR<c ==> α , $\alpha/(1-\beta) < c$.

- In the general case however LP and CP are contradictory.

Birnbaum’s general view of confidence

If there has been “one rock in a shifting scene” of general statistical thinking and practice in recent decades, it has not been the likelihood concept, as Edwards suggests, but rather the concept by which confidence limits and hypothesis tests are usually interpreted, which we may call the confidence concept of statistical evidence. This concept is not part of the Neyman-Pearson theory of tests and confidence region estimation, which denies any role to concepts of statistical evidence, as Neyman consistently insists. The confidence concept takes from the Neyman-Pearson approach techniques for systematically appraising and bounding the probabilities (under respective hypotheses) of seriously misleading interpretations of data. (The absence of a comparable property in the likelihood and Bayesian approaches is widely regarded as a decisive inadequacy.) The confidence concept also incorporates important but limited aspects of the likelihood concept: the sufficiency concept, expressed in the general refusal to use randomized tests and confidence limits when they are recommended by the Neyman-Pearson approach; and some applications of the conditionality concept. It is remarkable that this concept, an incompletely formalized synthesis of ingredients borrowed from mutually incompatible theoretical approaches, is evidently useful continuously in much critically informed statistical thinking and practice

Birnbaum , NATURE VOL. 225 MARCH 14, 1970 (1033)