Theory of Statistical Inference - Lecture III.6
STA422 and STA2162

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/sta422/sta4222026.html

2026

- the issue of properly defining statistical evidence has been a problem for the subject of statistical inference almost from its inception by Fisher

- Alan Birnbaum approached the problem by trying to elucidate various principles that serve to characterize the implications of a definition of statistical evidence, without formally defining the concept and established a theorem which generated a considerable amount of controversy over the years because its conclusion seems so counter-intuitive

- actually Birnbaum's treatment of the principles was a bit informal and, as we will see, this led to a misunderstanding of what his theorem says which removes the controversy to a great extent

- the formal treatment is as follows where we restrict the discussion to contexts where the sample space $\mathcal{X}$ and the model parameter space Θ are both finite as this is a rich enough environment to demonstrate all the relevant issues

- consider the set $\mathcal{I}$ containing all inference bases of the form $(\{f_\theta : \theta \in \Theta\}, x) = (M, x)$

- a *statistical principle* is then a relation defined on $\mathcal{I}$, namely, a statistical principle is a subset $\mathcal{R} \subset \mathcal{I} \times \mathcal{I}$

- Birnbaum's idea is that, if $(I_1, I_2) \in \mathcal{R}$, then the inference bases I_1, I_2 contain the same amount of evidence (sometimes called information) about the true value of the model parameter so all inferences **must** be the same for both inference bases

- as such, we want $\mathcal{R}$ to be an equivalence relation on $\mathcal{I}$
(i) reflexive - $(I, I) \in \mathcal{R}$ for all $I \in \mathcal{I}$
(ii) symmetric - if $(I_1, I_2) \in \mathcal{R}$, then $(I_2, I_1) \in \mathcal{R}$
(iii) transitive if $(I_1, I_2) \in \mathcal{R}, (I_2, I_3) \in \mathcal{R}$, then $(I_1, I_3) \in \mathcal{R}$

Definition III.6.1 *Invariance principle (parameter space)*: if

$$(\{f_\theta : \theta \in \Theta\}, x) = \left(\{f_\psi : \psi \in \Psi(\Theta)\}, x\right),$$

where $\Psi : \Theta \overset{1-1}{\to} \Psi(\Theta)$ and $\psi = \Psi(\theta)$, then these inferences bases are equivalent. ■

Exercise III.6.1 Show that the invariance principle is an equivalence relation on $\mathcal{I}$.

- this seems obvious but it is violated in the continuous case by some approaches (later)

- based on the invariance principle, and the fact that we are restricting the discussion to finite Θ, we will assume $\Theta = \{1, \ldots, k\}$ when $\#(\Theta) = k$

- there is an obvious *Invariance principle (sample space)* as well and we will assume these invariance principles hereafter

Definition III.6.2 *Likelihood principle* $\mathcal{L} : (I_1, I_2) \in \mathcal{L}$ whenever I_1 and I_2 give rise to proportional likelihoods. ■

Exercise III.6.2 Show that the likelihood principle is an equivalence relation on $\mathcal{I}$.

note - this is controversial because very different models can give rise to proportionate likelihoods for some x but repeated sampling properties are quite different

Example III.6.1

- suppose $I_1 = (\{X_{i=1}^n \text{Bernoulli}(\theta) : \theta \in (0,1)\}, x_1)$ where $x_1 \in \{0,1\}^n$ and $I_2 = (\{\text{negative binomial}(k,\theta) : \theta \in (0,1)\}, x_2)$ where $x_2 \in \mathbb{N}$ (# of tails until k-th head)

- then likelihoods are

$$L_1(\theta \,|\, x_1) \propto \theta^{\#\text{ of 1's}}(1-\theta)^{\#\text{ of 0's}} \text{ and } L_2(\theta \,|\, x_2) \propto \theta^k (1-\theta)^{x_2}$$

and so these are proportional when $(\#\text{ of 1's}) = k$ and $(\#\text{ of 0's}) = x_2$

- clearly repeated sampling properties are quite different ■

- later we'll see an approach that allows the inferences to be the same but the reliability assessments of the inferences (the real purpose of repeated sampling ideas) are quite different

- for model M with mss T, let M_T denote the marginal model for T call $(M_T, T(x))$ the minimal sufficient inference base derived from (M, x)

- recall if T, T' are mss's for M then there exists 1-1 function h s.t. $T' = h(T)$ and so by the invariance principle (sample space) $(M_T, T(x))$ and $(M_{T'}, T'(x))$ are equivalent

Definition III.6.3 *Sufficiency principle* $\mathcal{S}$: $(I_1, I_2) \in \mathcal{S}$ whenever I_1 and I_2 give rise to equivalent minimal sufficient inference bases. ■

Exercise III.6.3 Show that the sufficiency principle is an equivalence relation on $\mathcal{I}$.

note - clearly $\mathcal{S} \subset \mathcal{L}$ but note the model is part of $\mathcal{S}$ (via M_T) but is not part of $\mathcal{L}$

Exercise III.6.4 In Example III.6.1 derive the mss's and their models.

- what about the Conditionality principle?

- if A is ancillary for model $M = \{f_\theta : \theta \in \Theta\}$, let $M_{|A(x)} = \{f_\theta(\cdot \,|\, A(x)) : \theta \in \Theta\}$ denote the conditional model obtained by conditioning each distribution on $A(x)$ having occurred

Definition III.6.4 *Conditionality principle* $\mathcal{C}$: $(I_1, I_2) \in \mathcal{C}$ whenever I_2 is obtained from I_1 by conditioning on an ancillary for I_1 or conversely.

- clearly $\mathcal{C} \subset \mathcal{L}$ and so $\mathcal{S} \cup \mathcal{C} \subset \mathcal{L}$

note - both $\mathcal{S}$ and $\mathcal{C}$ are repeated sampling principles as the models are part of their specification

Birnbaum's Theorem as commonly stated: if a statistician accepts both $\mathcal{S}$ and $\mathcal{C}$ as basic principles of inference, then the statistician must accept $\mathcal{L}$.

- problem - there is no repeated sampling assessment allowed by $\mathcal{L}$

- this appears to be very paradoxical, what is going on?

- Evans (2020) What does the proof of Birnbaum's theorem prove? explains this (at least for me)

Lemma III.6.1 Suppose $\mathcal{R}$ is a reflexive relation on a set D. The the smallest equivalence relation containing $\mathcal{R}$ is given by

$$\begin{aligned} \bar{\mathcal{R}} &= \{(x, y) \in D^2 : \text{ for some } n \text{ there exists } x_1, \ldots, x_n \in D \text{ s.t.} \\ x &= x_1, y = x_n \text{ and } (x_i, x_{i+1}) \in D \text{ or } (x_{i+1}, x_i) \in D \text{ for all } i\}. \end{aligned}$$

Proof: Clearly $\bar{\mathcal{R}}$ is reflexive and symmetric. By construction if $(x, y) \in \bar{\mathcal{R}}$ and $(y, z) \in \bar{\mathcal{R}}$, then $(x, z) \in \bar{\mathcal{R}}$ (concatenate the chains) so $\bar{\mathcal{R}}$ is transitive and so is an equivalence relation. **Exercise III.6.5** Show that the intersection of any number of equivalence relations containing $\mathcal{R}$ is an equivalence relation. Since each such equivalence relation contains $\mathcal{R}$ it must be smallest eq. rel. that does so. ■

- given a relation $\mathcal{R}$ is $\bar{\mathcal{R}}$ meaningful, necessarily?

Example III.6.2

- let $D = \{2, 3, 4, \ldots\}$ and $\mathcal{R}$ be the relation $(x, y) \in \mathcal{R}$ when x and y have a common factor bigger than 1

- so $\mathcal{R}$ is reflexive and symmetric but not transitive as $(3, 6), (6, 8)$ shows

- but if $x, y \in D$, then $(x, xy), (xy, y) \in \mathcal{R}$ and so $(x, y) \in \bar{\mathcal{R}} = D^2$ and $\bar{\mathcal{R}}$ has no real meaning ■

- so the completion of a relation to be an eq. rel. is not necessarily meaningful and requires justification

- Birnbaum's Theorem properly stated

Theorem III.6.1 (*Birnbaum's Theorem*) $\mathcal{L} = \overline{\mathcal{S} \cup \mathcal{C}}$.
Proof: Suppose that $(I_1, I_2) \in \mathcal{L}$ where $I_i = (M_i, x_i)$. We build a chain from I_1 to I_2 using $\mathcal{S}$ and $\mathcal{C}$. Construct a new inference base $I = (M, y)$ from I_1 and I_2 as follows. Let M be given by
$\mathcal{X}_M = (\{1\} \times \mathcal{X}_{M_1}) \cup (\{2\} \times \mathcal{X}_{M_2})$,

$$f_{M,\theta}(1, x) = \begin{cases} (1/2) f_{M_1,\theta}(x) & \text{when } x \in \mathcal{X}_{M_1} \\ 0 & \text{otherwise,} \end{cases}$$
$$f_{M,\theta}(2, x) = \begin{cases} (1/2) f_{M_2,\theta}(x) & \text{when } x \in \mathcal{X}_{M_2} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$T(i, x) = \begin{cases} (i, x) & \text{when } x \notin \{x_1, x_2\} \\ \{x_1, x_2\} & \text{otherwise} \end{cases}$$

is sufficient for M and so $((M, (1, x_1)), (M, (2, x_2))) \in \mathcal{S}$.

Also, $h(i,x) = i$ is ancillary for M and thus

$$((M,(1,x_1)),(M_1,x_1)) \in \mathcal{C}, ((M,(2,x_2)),(M_2,x_2)) \in \mathcal{C}.$$

By Lemma III.6.1 $((M_1,x_1),(M_2,x_2)) \in \overline{\mathcal{S} \cup \mathcal{C}}$ and so $\mathcal{L} \subset \overline{\mathcal{S} \cup \mathcal{C}}$. But since $\mathcal{S} \cup \mathcal{C} \subset \mathcal{L}$ and $\mathcal{L}$ is an eq. rel. we have $\overline{\mathcal{S} \cup \mathcal{C}} \subset \mathcal{L}$ which completes the proof. ■

- this only has content if $\mathcal{S} \cup \mathcal{C} \neq \mathcal{L}$

Theorem III.6.2 $\mathcal{S} \cup \mathcal{C} \neq \mathcal{L}$
Proof: Suppose that M_1 has $\mathcal{X}_1 = \{0,1\}, \Theta_1 = \{1/5, 1/3\}$ with $f_{M_1,\theta}(x) = \theta^x(1-\theta)^{1-x}$ and M_2 has $\mathcal{X}_2 = \{0,1,2\}, \Theta_2 = \{1/5, 1/3\}$ with $f_{M_2,\theta}(0) = \theta, f_{M_2,\theta}(1) = \theta(1-\theta)$ and $f_{M_2,\theta}(2) = (1-\theta)^2$. Suppose further that $x_1 = 1$ and $x_2 = 0$ are observed so $f_{M_1,\theta}(1) = \theta = f_{M_2,\theta}(0)$. Note that the full data is minimal sufficient for both M_1 and M_2 and that both of these models have only trivial ancillaries. Therefore, if $I_1 = (M_1, 1)$ and $I_2 = (M_2, 0)$, then $(I_1, I_2) \notin S, (I_1, I_2) \notin C$ but $(I_1, I_2) \in L$ which proves that $\mathcal{S} \cup \mathcal{C}$ is properly contained in $\mathcal{L}$. ■

Exercise III.6.6 Verify the statements in the proof of Theorem III.6.2 concerning minimal sufficiency and ancillarity.

- in Example III.5.2 we had an inference base (M, x) that was related to two different inference bases $(M_{|A_1(x)}, x), (M_{|A_2(x)}, x)$ via $\mathcal{C}$, namely,

$$\Big((M, x), (M_{|A_1(x)}, x)\Big) \in \mathcal{C} \text{ and } \Big((M, x), (M_{|A_2(x)}, x)\Big) \in \mathcal{C}$$
$$\text{but } \Big((M_{|A_1(x)}, x), (M_{|A_2(x)}, x)\Big) \notin \mathcal{C}$$

and so $\mathcal{C}$ is not an eq. rel. which immediately establishes

Theorem III.6.3 $\mathcal{C} \subset \mathcal{L}$ but $\mathcal{C} \neq \mathcal{L}$

- the following result can be found in Evans (2012) based on an argument in Evans, Fraser and Monette (1986)

Theorem III.6.4 $\bar{\mathcal{C}} = \mathcal{L}$

Personal Opinion

Both Theorem III.6.1 and Theorem III.6.4 involve finding the smallest eq. rel. containing a relation, but a more logical approach would be to replace the relations $\mathcal{S} \cup \mathcal{C}$ *and* $\mathcal{C}$ *with the largest eq. rel. contained within them, see Evans and Franjakis (2023). This results in excluding some ancillaries from consideration. The usual statement of Birnbaum's Theorem is not correct but these results are still interesting and relevant to the quest of obtaining a good definition of what is meant by statistical evidence. On the other hand Birnbaum's Theorem can't be quoted as support for either pure likelihood or Bayesian inference.*