Statistical Inference
Lecture Two

https://people.bath.ac.uk/masss/APTS/apts.html

Simon Shaw

University of Bath

APTS, 14-18 December 2020

Overview of Lecture Two

- In Lecture One we considered a number of statistical principle, including the Weak Indifference Principle, WIP: if $f_X(x \mid \theta) = f_X(x' \mid \theta)$ for all $\theta \in \Theta$ then $\text{Ev}(\mathcal{E}, x) = \text{Ev}(\mathcal{E}, x')$.
- In Lecture Two, we'll consider further principles and their implications.
- Weak Conditionality Principle, WCP: if $\mathcal{E}^*$ is the mixture of the experiments $\mathcal{E}_1$, $\mathcal{E}_2$ according to mixture probabilities p_1, $p_2 = 1 - p_1$. then $\text{Ev}\left(\mathcal{E}^*, (i, x_i)\right) = \text{Ev}(\mathcal{E}_i, x_i)$.
- Strong Likelihood Principle, SLP: if $f_{X_1}(x_1 \mid \theta) = c(x_1, x_2) f_{X_2}(x_2 \mid \theta)$, for some function $c > 0$ for all $\theta \in \Theta$ then $\text{Ev}(\mathcal{E}_1, x_1) = \text{Ev}(\mathcal{E}_2, x_2)$.
- Birnbaum's Theorem: $(\text{WIP} \wedge \text{WCP}) \leftrightarrow \text{SLP}$.
- Stopping Rule Principle, SRP: in a sequential experiment $\mathcal{E}^\tau$, $\text{Ev}\left(\mathcal{E}^\tau, (x_1, \ldots, x_n)\right)$ does not depend on the stopping rule τ.
- $\text{SLP} \rightarrow \text{SRP}$.

The Likelihood Principle

- Consider experiments $\mathcal{E}_i = \{\mathcal{X}_i, \Theta, f_{X_i}(x_i \,|\, \theta)\}$, $i = 1, 2, \ldots$, where the parameter space Θ is the same for each experiment.
- Let $p_1, p_2, \ldots$ be a set of known probabilities so that $p_i \geq 0$ and $\sum_i p_i = 1$.

Mixture experiment

The mixture $\mathcal{E}^*$ of the experiments $\mathcal{E}_1, \mathcal{E}_2, \ldots$ according to mixture probabilities $p_1, p_2, \ldots$ is the two-stage experiment

1. A random selection of one of the experiments: $\mathcal{E}_i$ is selected with probability p_i.
2. The experiment selected in stage 1. is performed.

Thus, each outcome of the experiment $\mathcal{E}^*$ is a pair (i, x_i), where $i = 1, 2, \ldots$ and $x_i \in \mathcal{X}_i$, and family of distributions

$$f^*((i, x_i) \,|\, \theta) \;=\; p_i f_{X_i}(x_i \,|\, \theta).$$

Principle 4: Weak Conditionality Principle, WCP

Let $\mathcal{E}^*$ be the mixture of the experiments $\mathcal{E}_1$, $\mathcal{E}_2$ according to mixture probabilities p_1, $p_2 = 1 - p_1$. Then $\mathrm{Ev}\left(\mathcal{E}^*, (i, x_i)\right) = \mathrm{Ev}(\mathcal{E}_i, x_i)$.

- The WCP says that inferences for θ depend only on the experiment performed and not which experiments could have been performed.
- Suppose that $\mathcal{E}_i$ is randomly chosen with probability p_i and x_i is observed.
- The WCP states that the same evidence about θ would have been obtained if it was decided non-randomly to perform $\mathcal{E}_i$ from the beginning and x_i is observed.

Principle 5: Strong Likelihood Principle, SLP

Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be two experiments which have the same parameter θ. If $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$ satisfy $f_{X_1}(x_1 \,|\, \theta) = c(x_1, x_2) f_{X_2}(x_2 \,|\, \theta)$, that is

$$L_{X_1}(\theta; x_1) \;=\; c(x_1, x_2) L_{X_2}(\theta; x_2)$$

for some function $c > 0$ for all $\theta \in \Theta$ then $\mathrm{Ev}(\mathcal{E}_1, x_1) = \mathrm{Ev}(\mathcal{E}_2, x_2)$.

- The SLP states that if two likelihood functions for the same parameter have the same shape, then the evidence is the same.
- A corollary of the SLP, obtained by setting $\mathcal{E}_1 = \mathcal{E}_2 = \mathcal{E}$, is that $\mathrm{Ev}(\mathcal{E}, x)$ should depend on $\mathcal{E}$ and x only through $L_X(\theta; x)$.

Many classical statistical procedures violate the SLP and the following result was something of the bombshell, when it first emerged in the 1960s. The following form is due to Birnbaum (1972) and Basu (1975)

Birnbaum's Theorem

$(\text{WIP} \wedge \text{WCP}) \leftrightarrow \text{SLP}$.

Proof

Both SLP $\rightarrow$ WIP and SLP $\rightarrow$ WCP are straightforward. The trick is to prove $(\text{WIP} \wedge \text{WCP}) \rightarrow \text{SLP}$.
Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be two experiments which have the same parameter, and suppose that $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$ satisfy $f_{X_1}(x_1 \,|\, \theta) = c(x_1, x_2) f_{X_2}(x_2 \,|\, \theta)$ where the function $c > 0$. As the value c is known (as the data has been observed) then consider the mixture experiment with $p_1 = 1/(1+c)$ and $p_2 = c/(1+c)$.

Proof continued

$$f^*((1, x_1) \,|\, \theta) \;=\; \frac{1}{1+c} f_{X_1}(x_1 \,|\, \theta) \;=\; \frac{c}{1+c} f_{X_2}(x_2 \,|\, \theta) \;=\; f^*((2, x_2) \,|\, \theta)$$

Then the WIP implies that

$$\mathrm{Ev}\,(\mathcal{E}^*, (1, x_1)) \;=\; \mathrm{Ev}\,(\mathcal{E}^*, (2, x_2)).$$

Applying the WCP to each side we infer that

$$\mathrm{Ev}(\mathcal{E}_1, x_1) \;=\; \mathrm{Ev}(\mathcal{E}_2, x_2),$$

as required. □

Thus, either I accept the SLP, or I explain which of the two principles, WIP and WCP, I refute. Methods, which include many classical procedures, which violate the SLP face exactly this challenge.

The Sufficiency Principle

- Recall the idea of sufficiency: if $S = s(X)$ is sufficient for θ then

$$f_X(x \,|\, \theta) \;=\; f_{X|S}(x \,|\, s, \theta) f_S(s \,|\, \theta)$$

where $f_{X|S}(x \,|\, s, \theta)$ does not depend upon θ.

- Consequently, consider the experiment $\mathcal{E}^S = \{s(\mathcal{X}), \Theta, f_S(s \,|\, \theta)\}$.

Principle 6: Strong Sufficiency Principle, SSP

If $S = s(X)$ is a sufficient statistic for $\mathcal{E} = \{\mathcal{X}, \Theta, f_X(x \,|\, \theta)\}$ then $\mathrm{Ev}(\mathcal{E}, x) = \mathrm{Ev}(\mathcal{E}^S, s(x))$.

Principle 7: Weak Sufficiency Principle, WSP

If $S = s(X)$ is a sufficient statistic for $\mathcal{E} = \{\mathcal{X}, \Theta, f_X(x \,|\, \theta)\}$ and $s(x) = s(x')$ then $\mathrm{Ev}(\mathcal{E}, x) = \mathrm{Ev}(\mathcal{E}, x')$.

Theorem

SLP $\rightarrow$ SSP $\rightarrow$ WSP $\rightarrow$ WIP.

Proof

As s is sufficient, $f_X(x \mid \theta) = cf_S(s \mid \theta)$ where $c = f_{X|S}(x \mid s, \theta)$ does not depend on θ. Applying the SLP, $\text{Ev}(\mathcal{E}, x) = \text{Ev}(\mathcal{E}^S, s(x))$ which is the SSP. Note, that from the SSP,

$$\begin{aligned}
\text{Ev}(\mathcal{E}, x) &= \text{Ev}(\mathcal{E}^S, s(x)) && \text{(by the SSP)} \\
&= \text{Ev}(\mathcal{E}^S, s(x')) && (\text{as } s(x) = s(x')) \\
&= \text{Ev}(\mathcal{E}, x') && \text{(by the SSP)}
\end{aligned}$$

We thus have the WSP. Finally, if $f_X(x \mid \theta) = f_X(x' \mid \theta)$ as in the statement of WIP then $s(x) = x'$ is sufficient for x. Hence, from the WSP, $\text{Ev}(\mathcal{E}, x) = \text{Ev}(\mathcal{E}, x')$ giving the WIP. $\square$

If we put together the last two theorems, we get the following corollary.

Corollary

$(\text{WIP} \wedge \text{WCP}) \rightarrow \text{SSP}$.

Proof

From Birnbaum's theorem, $(\text{WIP} \wedge \text{WCP}) \leftrightarrow \text{SLP}$ and from the previous theorem, $\text{SLP} \rightarrow \text{SSP}$. $\square$

- Birnbaum's (1962) original result combined sufficiency and conditionality for the likelihood but he revised this to the WIP and WCP in later work.
- One advantage of this is that it reduces the dependency on sufficiency: Pitman-Koopman-Darmois Theorem states that sufficiency more-or-less characterises the exponential family.

Stopping rules

- Consider observing a sequence of random variables $X_1, X_2, \ldots$ where the number of observations is not fixed in advance but depends on the values seen so far.
  - At time j, the decision to observe X_{j+1} can be modelled by a probability $p_j(x_1, \ldots, x_j)$.
  - We assume, resources being finite, that the experiment must stop at specified time m, if it has not stopped already, hence $p_m(x_1, \ldots, x_m) = 0$.
- The stopping rule may then be denoted as $\tau = (p_1, \ldots, p_m)$. This gives an experiment $\mathcal{E}^{\tau}$ with, for $n = 1, 2, \ldots$, $f_n(x_1, \ldots, x_n \,|\, \theta)$ where consistency requires that

$$f_n(x_1, \ldots, x_n \,|\, \theta) \;=\; \sum_{x_{n+1}} \cdots \sum_{x_m} f_m(x_1, \ldots, x_n, x_{n+1}, \ldots x_m \,|\, \theta).$$

Motivation for the stopping rule principle (Basu, 1975)

- Consider four different coin-tossing experiments (with some finite limit on the number of tosses).
  - $\mathcal{E}_1$ Toss the coin exactly 10 times;
  - $\mathcal{E}_2$ Continue tossing until 6 heads appear;
  - $\mathcal{E}_3$ Continue tossing until 3 consecutive heads appear;
  - $\mathcal{E}_4$ Continue tossing until the accumulated number of heads exceeds that of tails by exactly 2.
- Suppose that all four experiments have the same outcome $x = (\mathrm{T,H,T,T,H,H,T,H,H,H})$.
- We may feel that the evidence for θ, the probability of heads, is the same in every case.
  - Once the sequence of heads and tails is known, the intentions of the original experimenter (i.e. the experiment she was doing) are immaterial to inference about the probability of heads.
  - The simplest experiment $\mathcal{E}_1$ can be used for inference.

Principle 8: Stopping Rule Principle, SRP

[a] In a sequential experiment $\mathcal{E}^{\tau}$, $\text{Ev}\left(\mathcal{E}^{\tau}, (x_1, \ldots, x_n)\right)$ does not depend on the stopping rule τ.

[a]Basu (1975) claims the SRP is due to George Barnard (1915-2002)

- If it is accepted, the SRP is nothing short of revolutionary.
- It implies that the intentions of the experimenter, represented by τ, are irrelevant for making inferences about θ, once the observations $(x_1, \ldots, x_n)$ are known.
- Once the data is observed, we can ignore the sampling plan.
- The statistician could proceed as though the simplest possible stopping rule were in effect, which is $p_1 = \cdots = p_{n-1} = 1$ and $p_n = 0$, an experiment with n fixed in advance, $\mathcal{E}^n = \{\mathcal{X}_{1:n}, \Theta, f_n(x_{1:n} \mid \theta)\}$.
- Can the SRP possibly be justified? Indeed it can.

Theorem

SLP $\to$ SRP.

Proof

Let τ be an arbitrary stopping rule, and consider the outcome $(x_1, \ldots, x_n)$, which we will denote as $x_{1:n}$.

- We take the first observation with probability one.
- For $j = 1, \ldots, n-1$, the $(j+1)$th observation is taken with probability $p_j(x_{1:j})$.
- We stop after the nth observation with probability $1 - p_n(x_{1:n})$.

Consequently, the probability of this outcome under τ is

$$f_\tau(x_{1:n} \,|\, \theta) \;=\; f_1(x_1 \,|\, \theta) \left\{ \prod_{j=1}^{n-1} p_j(x_{1:j}) \, f_{j+1}(x_{j+1} \,|\, x_{1:j}, \theta) \right\} (1 - p_n(x_{1:n}))$$

Proof continued

$$
\begin{aligned}
f_\tau(x_{1:n} \mid \theta) &= \left\{\prod_{j=1}^{n-1} p_j(x_{1:j})\right\}(1 - p_n(x_{1:n}))\, f_1(x_1 \mid \theta) \prod_{j=2}^{n} f_j(x_j \mid x_{1:(j-1)}, \theta) \\
&= \left\{\prod_{j=1}^{n-1} p_j(x_{1:j})\right\}(1 - p_n(x_{1:n})) f_n(x_{1:n} \mid \theta).
\end{aligned}
$$

Now observe that this equation has the form

$$
f_\tau(x_{1:n} \mid \theta) = c(x_{1:n}) f_n(x_{1:n} \mid \theta) \qquad (1)
$$

where $c(x_{1:n}) > 0$. Thus the SLP implies that $\mathrm{Ev}(\mathcal{E}^\tau, x_{1:n}) = \mathrm{Ev}(\mathcal{E}^n, x_{1:n})$ where $\mathcal{E}^n = \{\mathcal{X}_{1:n}, \Theta, f_n(x_{1:n} \mid \theta)\}$. Since the choice of stopping rule was arbitrary, equation (1) holds for all stopping rules, showing that the choice of stopping rule is irrelevant. $\square$

A comment from Leonard Jimmie Savage (1917-1971), one of the great statisticians of the Twentieth Century, captured the revolutionary and transformative nature of the SRP.

> *May I digress to say publicly that I learned the stopping rule principle from Professor Barnard, in conversation in the summer of 1952. Frankly, I then thought it a scandal that anyone in the profession could advance an idea so patently wrong, even as today I can scarcely believe that some people resist an idea so patently right. (Savage et al., 1962, p76)*