Statistical Inference
Lecture One

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

Simon Shaw

University of Bath

APTS, 14-18 December 2020

Overview of Lecture One

- We wish to consider inferences about a parameter θ given a parametric model $\mathcal{E} = \{\mathcal{X}, \Theta, f_X(x \,|\, \theta)\}$

$$(\mathcal{E}, x) \xmapsto{\text{statistician, Ev}} \text{Inference about } \theta.$$

- We'll consider a series of statistical principles to guide the way to learn about θ.
- Weak Indifference Principle, WIP: if $f_X(x \,|\, \theta) = f_X(x' \,|\, \theta)$ for all $\theta \in \Theta$ then $\text{Ev}(\mathcal{E}, x) = \text{Ev}(\mathcal{E}, x')$.
- Distribution Principle, DP: if $\mathcal{E} = \mathcal{E}'$, then $\text{Ev}(\mathcal{E}, x) = \text{Ev}(\mathcal{E}', x)$.
- Transformation Principle, TP: for the bijective $g : \mathcal{X} \to \mathcal{Y}$, construct $\mathcal{E}^g = \{\mathcal{Y}, \Theta, f_Y(y \,|\, \theta)\}$. Then $\text{Ev}(\mathcal{E}, x) = \text{Ev}(\mathcal{E}^g, g(x))$.
- $(\text{DP} \wedge \text{TP}) \to \text{WIP}$.

Introduction

- We wish to consider inferences about a parameter θ given a parametric model

$$\mathcal{E} \; = \; \{\mathcal{X}, \Theta, f_X(x \,|\, \theta)\}.$$

- We assume that the model is true so that only $\theta \in \Theta$ is unknown. We wish to learn about θ from observations x (typically, vector valued) so that $\mathcal{E}$ represents a model for this experiment.

Smith (2010) considers that there are three players in an inference problem:

1. Client: person with the problem
2. Statistician: employed by the client to help solve the problem
3. Auditor: hired by the client to check the statistician's work

The statistician is thus responsible for explaining the rationale behind the choice of inference in a compelling way.

Reasoning about inferences

We consider a series of statistical principles to guide the way to learn about θ. The principles are meant to be either self-evident or logical implications of principles which are self-evident.
We shall assume that $\mathcal{X}$ is finite: Basu (1975) argues that "infinite and continuous models are to be looked upon as mere approximations to the finite realities."

- Inspiration of Allan Birnbaum (1923-1976) to see how to construct and reason about statistical principles given "evidence" from data.
- The model $\mathcal{E} = \{\mathcal{X}, \Theta, f_X(x \mid \theta)\}$ is accepted as a working hypothesis.
- How the statistician chooses her inference statements about the true value θ is entirely down to her and her client.
  - as a point or a set in Θ;
  - as a choice among alternative sets or actions;
  - or maybe as something more complicated, not ruling out visualisations.

- Following Dawid (1977), consider that the statistician defines, a priori, a set of possible inferences about θ
- Task is to choose an element of this set based on $\mathcal{E}$ and x.
- The statistician should see herself as a function Ev: a mapping from $(\mathcal{E}, x)$ into a predefined set of inferences about θ.

$$(\mathcal{E}, x) \xmapsto{\text{statistician, Ev}} \text{Inference about } \theta.$$

- For example, $\text{Ev}(\mathcal{E}, x)$ might be:
  - the maximum likelihood estimator of θ
  - a 95% confidence interval for θ
- Birnbaum called $\mathcal{E}$ the experiment, x the outcome, and Ev the evidence.

Note:

1. There can be different experiments with the same θ.
2. Under some outcomes, we would agree that it is self-evident that these different experiments provide the same evidence about θ.

Example

Consider two experiments with the same θ.

1. $X \sim Bin(n, \theta)$, so we observe x successes in n trials.
2. $Y \sim NBin(r, \theta)$, so we observe the rth success in the yth trial.

If we observe $x = r$ and $y = n$, do we make the same inference about θ in each case?

Consider two experiments $\mathcal{E}_1 = \{\mathcal{X}_1, \Theta, f_{X_1}(x_1 \,|\, \theta)\}$ and $\mathcal{E}_2 = \{\mathcal{X}_2, \Theta, f_{X_2}(x_2 \,|\, \theta)\}$.

Equivalence of evidence (Basu, 1975)

The equality or equivalence of $\text{Ev}(\mathcal{E}_1, x_1)$ and $\text{Ev}(\mathcal{E}_2, x_2)$ means that:

1. $\mathcal{E}_1$ and $\mathcal{E}_2$ are related to the same parameter θ.
2. Everything else being equal, the outcome x_1 from $\mathcal{E}_1$ warrants the same inference about θ as does the outcomes x_2 from $\mathcal{E}_2$.

- We now consider constructing statistical principles and demonstrate how these principles imply other principles.
- These principles all have the same form: under such and such conditions, the evidence about θ should be the same.
- Thus they serve only to rule out inferences that satisfy the conditions but have different evidences. They do not tell us how to do an inference, only what to avoid.

The principle of indifference

Principle 1: Weak Indifference Principle, WIP

Let $\mathcal{E} = \{\mathcal{X}, \Theta, f_X(x \,|\, \theta)\}$. If $f_X(x \,|\, \theta) = f_X(x' \,|\, \theta)$ for all $\theta \in \Theta$ then $\text{Ev}(\mathcal{E}, x) = \text{Ev}(\mathcal{E}, x')$.

- We are indifferent between two models of evidence if they differ only in the manner of the labelling of sample points.
- If $X = (X_1, \ldots, X_n)$ where the X_is are a series of independent Bernoulli trials with parameter θ then $f_X(x \,|\, \theta) = f_X(x' \,|\, \theta)$ if x and x' contain the same number of successes.

Principle 2: Distribution Principle, DP

If $\mathcal{E} = \mathcal{E}'$, then $\text{Ev}(\mathcal{E}, x) = \text{Ev}(\mathcal{E}', x)$.

- Informally, (Dawid, 1977), only aspects of an experiment which are relevant to inference are the sample space and the family of distributions over it.

Principle 3: Transformation Principle, TP

Let $\mathcal{E} = \{\mathcal{X}, \Theta, f_X(x \,|\, \theta)\}$. For the bijective $g : \mathcal{X} \to \mathcal{Y}$, let $\mathcal{E}^g = \{\mathcal{Y}, \Theta, f_Y(y \,|\, \theta)\}$, the same experiment as $\mathcal{E}$ but expressed in terms of $Y = g(X)$, rather than X. Then $\text{Ev}(\mathcal{E}, x) = \text{Ev}(\mathcal{E}^g, g(x))$.

- Inferences should not depend on the way in which the sample space is labelled, for example, X or X^{-1}.

Theorem

(DP $\wedge$ TP) $\rightarrow$ WIP.

Proof

Fix $\mathcal{E}$, and suppose that $x, x' \in \mathcal{X}$ satisfy $f_X(x \,|\, \theta) = f_X(x' \,|\, \theta)$ for all $\theta \in \Theta$, as in the condition of the WIP.
Let $g : \mathcal{X} \rightarrow \mathcal{X}$ be the function which switches x for x', but leaves all of the other elements of $\mathcal{X}$ unchanged. Then $\mathcal{E} = \mathcal{E}^g$ and

$$\begin{aligned} \mathrm{Ev}(\mathcal{E}, x') &= \mathrm{Ev}(\mathcal{E}^g, x') \quad \text{[by the DP]} \\ &= \mathrm{Ev}(\mathcal{E}^g, g(x)) \\ &= \mathrm{Ev}(\mathcal{E}, x), \quad \text{[by the TP]} \end{aligned}$$

which gives the WIP. □