Induction Conference
Ghent 2008

What is a Statistical Hypothesis?

⋆

Jan-Willem Romeijn

Faculty of Philosophy
University of Groningen

Contents

1 What this talk is about

Statistical hypotheses play an important role in science, especially in the behavioural and social sciences.

Psychology: Latent psychological attributes are tied to psychological test by statistical relations.

Sociology: Variation due to individual behaviour is captured by describing groups in statistical terms.

Economics: Trends are described by time-series that include a systematic component and a statistical error.

What are statistical hypotheses?

Statistical hypotheses are probability distributions. But how can we interpret and use these distributions?

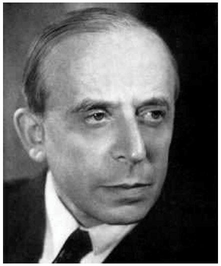
- How do probability distributions relate to empirical fact?
- What is the role of statistical hypotheses in inference?

In response to this, I argue that

- statistical hypotheses have a specific empirical content related to frequentism, and that
- as such they have a definite function in the formal semantics of statistical inference.

2 Von Mises' frequentism

An important starting point in the discussion on statistics is the frequentist interpretation of probability by von Mises.



It is based on the empirical notion of a collective, defined by a limiting relative frequency and excluded gambling systems.

Problems with frequentism

Frequentism has been criticised heavily.

- ↯ Infinite sequences: in the long run, we are all dead (Keynes).
- ↯ Random sequences: seemingly intentional aspects to place selections (Ville, but see van Lambalgen).
- ↯ Finite frequentism: a host of problems (Hajek).

I do not want to claim that von Mises cannot defend his theory against all these criticisms. Still...

Reversed frequentism

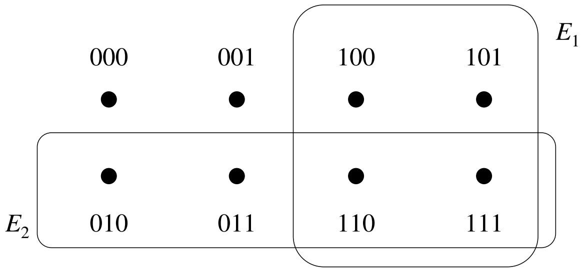
In this talk I develop a different view. I argue for the following claims:

- Frequentism is best seen as a formal semantics for statistical inference, not as a physical theory of mass phenomena.
- Statistical hypotheses have empirical content, namely of frequency in the limit.

It will be seen that this has distinct advantages for understanding the role of statistical hypotheses in inference.

3 Statistical hypotheses in sample space

A statistical analysis is always based on a set of possible observations, a sample space. For tossing a coin N times, the sample space is $\{0, 1\}^N$.



Samples e_t can be represented as sets E_t in this space.

Hypotheses as distributions

We may also construct an idealised sample space consisting of infinitely long samples: $\{0,1\}^{\Omega}$.

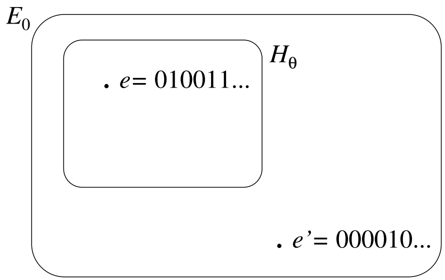
$$e = 010011011001010\ldots$$

$$p_{\theta}(E_t|E_{t'}) = \theta$$

We can define the statistical hypothesis h_{θ} as a distribution over this infinite sample space.

Hypotheses as tail events

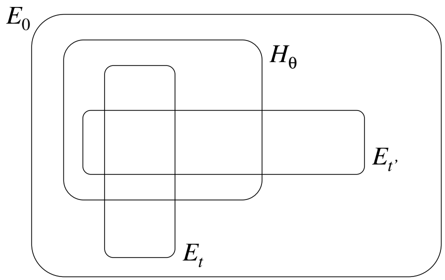
Some elements $e = 00110111...$ of this sample space are collectives in the sense of von Mises.



We can identify statistical hypotheses h_θ with the set of all collectives H_θ that, according to frequentism, instantiate the probability distribution p_θ.

Tail events as distributions

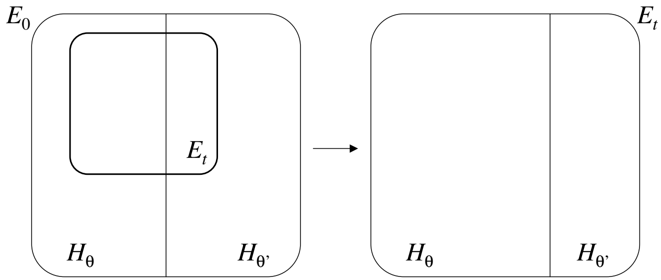
Each set H_θ intersects with all sets E_t that are assigned nonzero probability by p_θ.



We can link sets of collectives and distributions by means of a kind of ergodicity assumption.

4 Statistical hypotheses in inference

The principle of Bayesian inference is conditionalisation. Data e_t are reflected in the probability assignment by zooming in on the probability assignment within e_t: $p(E_{t'}) \rightarrow p(E_{t'}|E_t)$.



We can now apply this idea to hypotheses as well.

Hypotheses fix inductive dependence

In Carnapian inductive logic, the inductive influence of observations e_t on expectations over other observations $e_{t'}$ is fixed directly, by means of inductive rules.

A set of statistical hypotheses, or a model for short, provides a convenient way of fixing this inductive dependency; hypotheses extend the language of inductive logic.

5 What is the use?

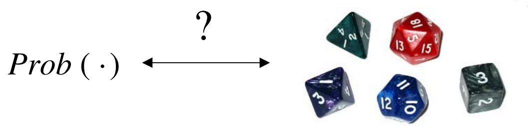
The aim was to explain the relation between statistical hypotheses and empirical fact, and the role of these hypotheses in statistical inference.

- The statistical hypotheses are given empirical content in the idealised setting of an infinite sample space.
- As such, the hypotheses perform a definite function in a logic of statistical inference.

So by putting statistical hypotheses in the above perspective, we provide these explanations.

Employing frequentism

Von Mises presented frequentism as a theory on what probability is, empirically grounding it in mass phenomena.



Instead, frequentism is here used to interpret statistical hypotheses given beforehand. A notion of probability is presupposed, and frequentism is used to relate it to empirical fact.

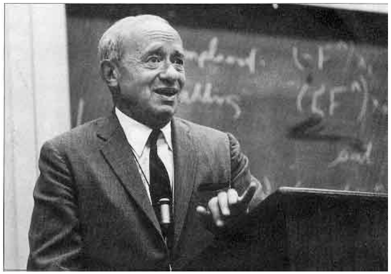
Formal semantics

The problematic notion of collective is taken as part of a formal semantics, and thus of model theory. Therefore many of the criticisms against von Mises do not apply.

In its formal semantical role, we can find a place for frequentism within an epistemic, either subjective or objective, interpretation of probability.

Inductive logic

The success of deductive logic is partly based on a clear separation of syntax and semantics, which led to notions such as validity, truth in the model, etc.



I hope the formal semantics are a step towards putting inductive logic on the same footing as deductive logic.

6 Convergence and completeness

We have sketched a formal semantics for statistical inference. We now start on some model theory and logic.

Soundness: if the inference machinery proves a proposition, it is guaranteed to be true in the model.

Completeness: if a proposition is true in the model, the inference machinery is guaranteed to prove it.

Clearly we must assume that all premises used to prove propositions are true.

Completeness for statistical inference

For Bayesian statistical inference we can reformulate completeness as convergence.

Completeness: If a hypothesis is true, in the long run Bayesian statistical inference will give it probability 1.

Here we assume that we do not assign zero probability to true propositions, and that the data separate the hypotheses, meaning that the data can tell hypotheses apart in the limit.

The well-known convergence theorem by Gaifman and Snir may then be used as a completeness theorem.

Thanks!

This talk will be available at http://www.philos.rug.nl/ romeyn.
For comments and questions, email j.w.romeijn@rug.nl.