Epistemic Statistics

Jan-Willem Romeijn
University of Groningen
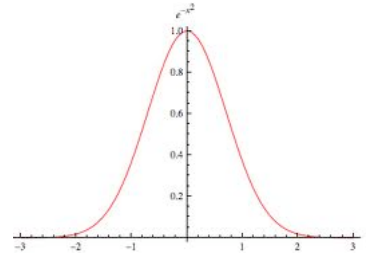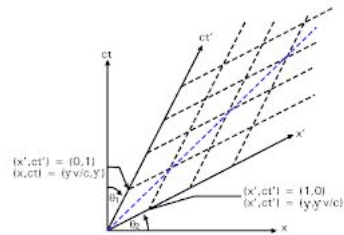Partly joint work with Richard Morey (RuG)

Richard Morey

Assistant professor in psychological methods with a keen interest in foundational matters.

Statistics is difficult

My first grade in physics was a 65% for the statistics course "Measuring in Physics 1". My second was a 95% on "Special Relativity".

Classical statistics seemed more difficult than special relativity. Why is this?
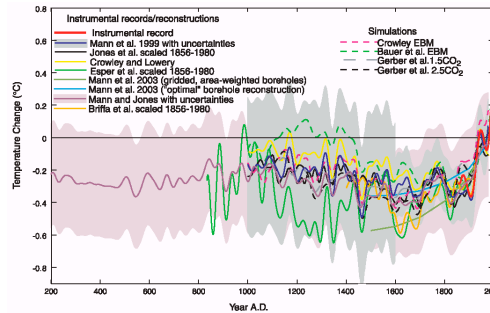
The trouble with statistics

One explanation is that the theory of statistics does not accord with its intuitive use.

Despite Fisher's resistance, classical statistics was explicitly designed not to be concerned with belief.

Science and belief

Statistics is at the very heart of science, and misunderstandings over what it tells us are rampant.

Bringing belief back in the picture will significantly improve science and the policies based on it.

Contents

1 Classical statistics

Classical statistical results are hard to interpret. We will look at three standard result types, to wit. . .

- the p-value of a Fisher null hypothesis test,

- power and significance in a Neyman-Pearson test, and

- the confidence interval around the estimation of a parameter value.

None of these allow for an interpretation in terms of belief concerning hypotheses.

The null hypothesis test

Consider orchards producing pears of three different colours. We sample one pear from a truck that came from Anna's orchard.

Hypothesis \ Data	Red	Green	Yellow
Anna	0.05	0.05	0.90

If we follow the rule to

reject that the truck came from Anna when drawing a red pear

then we go wrong in 5% of cases. The same goes for the green pear.

Neyman-Pearson testing

Comparison to an alternative hypothesis helps to determine which colour of pear is critical.

Hypothesis \ Data	Red	Green	Yellow
Anna	0.05	0.05	0.95
Ben	0.40	0.30	0.30

In this case red is critical. The optimal test rules out that the truck came from Anna with...

- 5% false rejections or *significance*, and

- 60% false acceptance, or 40% *power*.

A "remarkable procedure"

Compare testing Anna against Ben with testing Chris against him. A green pear licenses rejection of Chris as origin, and acceptance of Anna.

Hypothesis \ Data	Red	Green	Yellow
Anna	0.05	0.05	0.95
Ben	0.40	0.30	0.30
Chris	0.00	0.05	0.90

But if the truck did come from Chris, we falsely rejected this hypothesis because, in comparison to Anna,

"it fails to predict an outcome that does not occur" (cf. Jeffreys).

Confidence intervals

The 95% confidence interval, despite its epistemic ring, does not support an epistemic reading either.
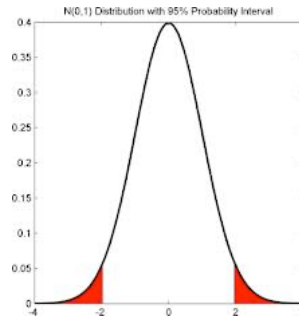
The interval designates the range of parameter values that do not make the data improbable. Nothing is said about confidence pertaining to parameter values.

2   Bayesian statistics

Bayesian statistics has more direct bearing on belief. Is all this mere window dressing for Bayesian propaganda?

- The foundations of Bayesian statistics are strongly associated with decision theory.

- Practicing Bayesian statisticians disassociate themselves from epistemic interpretations.

The Bayesian formalism accommodates an epistemic interpretation, but both theory and practice stop short of it.

Opinion dynamics

Central to the formalism is the impact of evidence on the probability over statistical hypotheses.

Hypothesis \ Probability	Prior	Evidence	Posterior
Anna	0.5	0.05	0.14
Ben	0.5	0.30	0.86

Priors and posteriors range over hypotheses and seem directly related to belief.

Pragmatic and objective Bayesianism

The formalism is not often interpreted as pertaining to belief though.

- Priors are portrayed as neutral starting point, typically invoking the principle of indifference.

- Results are reported as likelihood ratios, avoiding talk of priors altogether.

Statisticians often use Bayesian methods pragmatically, without buying the epistemic interpretation.

Interpreting the probabilities

On closer inspection, this interpretation of probability in Bayesian statistics is not exactly epistemic either.

Central to the interpretation of probabilities is their use in rational decision and action, for instance via a semantics based on betting behavior.

Behaviourism and empiricism

The emphasis on decision and action chimes with behaviourism: only the observable consequences of belief matter.

This idea has driven the development of both classical and Bayesian statistics.

3 Science as epistemic project

We believe that the focus of statistical method on decision and action is detrimental to science.

- Currently statistical methods invite misapplication, and lead to faulty and idiosyncratic relations to epistemic attitudes.

- It is unrealistic to deny researchers access to epistemic parlance, and force them into an instrumentalist attitude.

Statisticians, philosophers, and researchers need to work together to revamp statistics.

Past epistemic statistics

Several attempts were made to retain an epistemic component for statistics.

- De Finetti emphasized the logical nature of statistical inference and the epistemic interpretation of its results.

- Fisher devised the fiducial argument, allowing us to derive an epistemic probability over hypotheses without employing a prior.

- Kyburg and co-workers developed a defeasible logic to match classical statistics.

These enjoy cult status among philosophers of statistics but they have not moved the masses.

Current epistemic statistics

Recent attempts to make the epistemic implications of statistics more visible:

- Kadane and Goldstein emphasize the radical subjectivity of all probability.

- According to Royall and Sober, the likelihoods of hypotheses express the strength of evidence.

- Mayo and Spanos revamp Neyman-Pearson statistics in terms of error avoidance and reliability.

Authors either grudgingly accept the epistemic and then downplay it, or else surrender to it completely.

Dig deeper

What are the deep roots of this preference for behaviourism, empiricism, objectivism?

Epistemological readings of statistics seem to invite relativism about our knowledge.

Hume's problem

It reveals a well-known problem for scientific method in general: the problem of induction.

To solve this problem for statistics, it seems that we must diminish its subject-relative component.

The Kantian response

I propose to turn the table: the input of the subject makes it possible that we learn from data.

Accepting that a method has a subject-relative component does not lead to full-blown relativism.

4 Inferential statistics

Statistical results are much like conclusions of deductive inference. They depend on premises about sample space, model, prior, and so on.
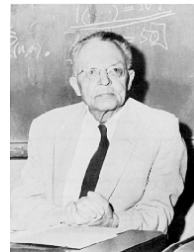
Statistics thus provides constraints on what to believe of a conditional nature.

No "free lunch" but a cheap one

The challenge is to arrive at reliable conclusions by minimizing, motivating, or justifying the premises.

The statistician's job is to ensure correct procedures, and to design ones that are conducive to this optimization goal.

Universal, local, or material induction

Importantly, there may be many logics of statistics. Induction may be local, or material. Its results need not be unique.

Nevertheless, viewing statistics as inferential will help giving it the appropriate role in science.

5 Statistics as applied philosophy

To connect statistical method and the epistemic aims of science all stakeholders have work to do.

Researchers need to communicate explicitly about the epistemic goals of their data collection and theorizing.

Statisticians must be clear on the epistemic status of their methods, and design new ones that fit the needs of researchers.

Philosophers have to realize the potential of their insights, and collaborate with researchers to realize this.

Thank you

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