Statistics and Uncertainty in High Energy Physics

Kent Staley

CERN History Days

February 1, 2018

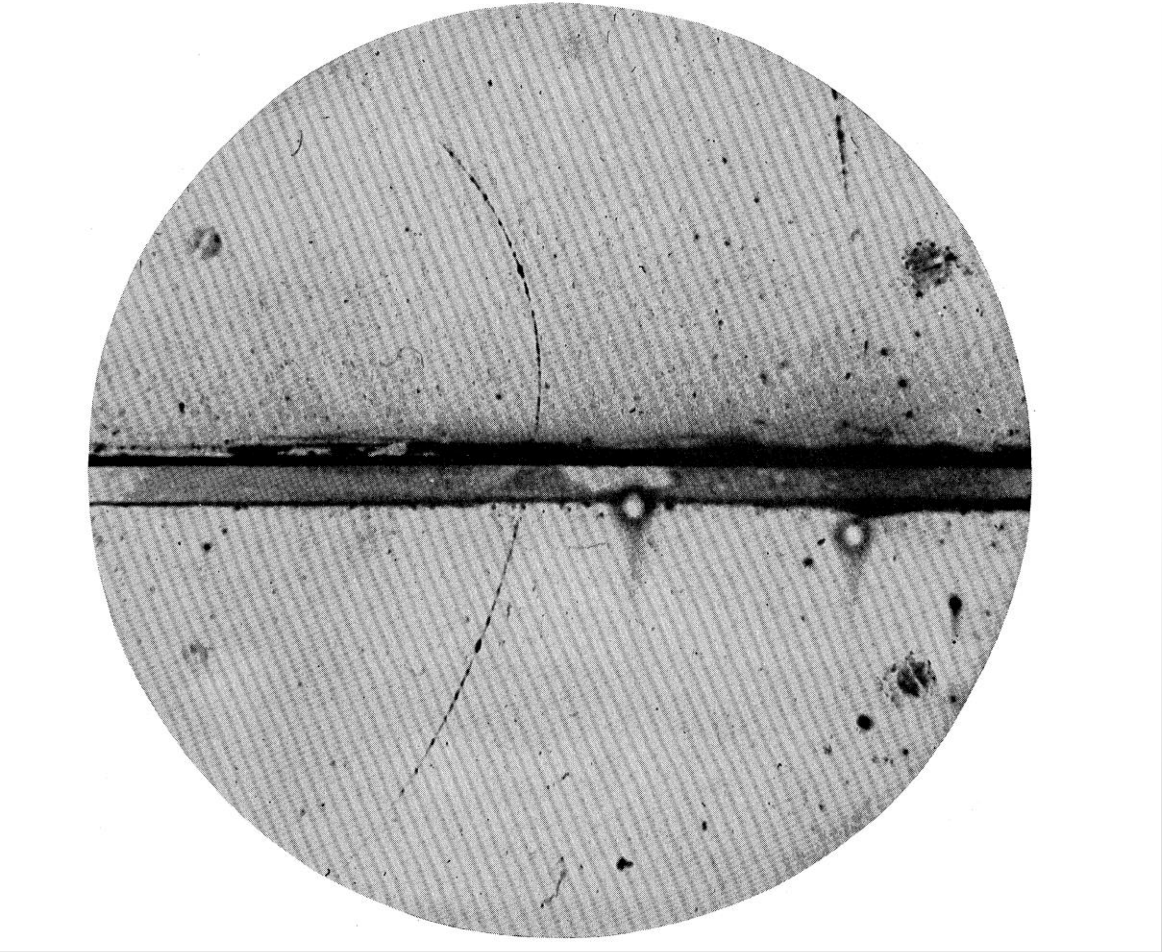
SAINT LOUIS UNIVERSITY

Introduction

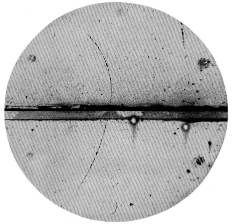
Aims of this talk

- Emphasize the essential role of statistical argumentation in experimental HEP
- Situate statistical practice in HEP relative to foundational debates regarding statistics
- Sketch a justification for HEP’s preference for frequentist over Bayesian statistics.
- Indicate challenges in the statistical characterization of systematic uncertainty calling for interdisciplinary efforts.

The indispensability of statistical argumentation



A “golden event” discovery with statistical argumentation



- Galison (1997): The image tradition allows discovery without statistics because a single image is sufficient.
- Anderson (1933): “[The background hypothesis of an accidental coincidence of two negative electrons] was dismissed on a probability basis, since a sharp track of this order of curvature under the experimental conditions prevailing occurred in the chamber only once in some 500 exposures, and since there was practically no chance at all that two such tracks should line up in this way (Anderson 1933, p. 491).

Frequentism and Bayesianism

Two probability frameworks: Frequentism

- “objective” – used to characterize regularities and variabilities of data-generating mechanisms
- idealized – probabilities correspond to “long run” relative frequencies
- implemented through a well-established statistical toolkit
- Inferences are conceived of as the outcome of testing/estimation procedures.
- The primary aim is to be able to characterize the severity of procedures resulting in a given inference.

Two probability frameworks: Bayesianism

- “subjective” – epistemic: used to characterize degrees of belief
- idealized – probabilities represent degrees of belief of a coherent epistemic agent
- utilization by scientists for purposes of analyzing experimental data has been less widespread
- central role of Bayes’ theorem:
  - $p(\mu\ /\ x_0) = [p(\mu)p(x_0)]\ /\ \int p(x_0\ /\ \mu)p(\mu)d\mu$
  - $p(H\ /\ E) = [p(H)p(E)]/p(E)$
- The primary aim is to determine a posterior probability distribution.

Pragmatic frequentism

Why frequentism?

- HEP relies primarily on frequentist statistics in the analysis of data.
- A preference for a Bayesian approach is widespread amongst philosophers of science, and is growing amongst statisticians.
- The use of p-values is subject to routine criticisms and calls for reform.
- Why do physicists continue to rely on frequentist methods generally and p-values in particular?

A pragmatist response

- The justification for use of frequentist methods is pragmatic: they are useful for meeting the practical demands of argumentation in pursuit of the scientific aims of HEP.
- This is in contrast to thinking of these methods as a device for determining personal or collective beliefs.

An assumption about the aims of experimental HEP

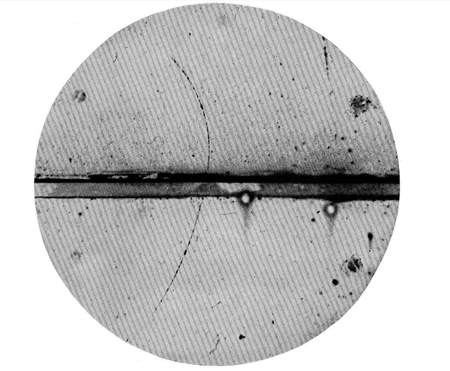
- HEP physicists aim not (only) to form beliefs about physical phenomena within the domain of HEP, but to collectively contribute to the production of shared scientific knowledge about those phenomena.
  - epistemic – knowledge-related
  - pragmatic – an activity (in which argumentation is central) undertaken collectively

Discovery is a practical task

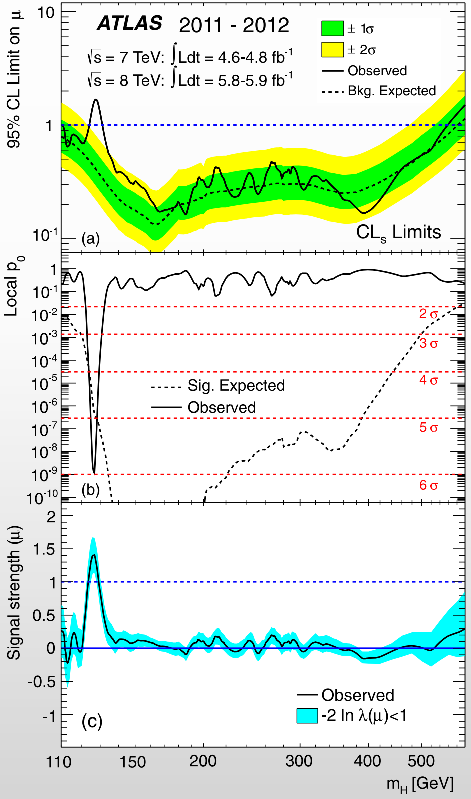
- To make a discovery claim is not merely to report a high posterior probability for a proposition asserting an existence claim.
- To discover something requires that one possess significant new evidence in support of such a claim.
- To argue that one possess such evidence requires establishing the incompatibility of that evidence with any background hypothesis (necessary but not sufficient).
- p-values are very well suited for this task.

From Anderson…

- Anderson (1933): “[The background hypothesis of an accidental coincidence of two negative electrons] was dismissed on a probability basis, since a sharp track of this order of curvature under the experimental conditions prevailing occurred in the chamber only once in some 500 exposures, and since there was practically no chance at all that two such tracks should line up in this way (Anderson 1933, p. 491).

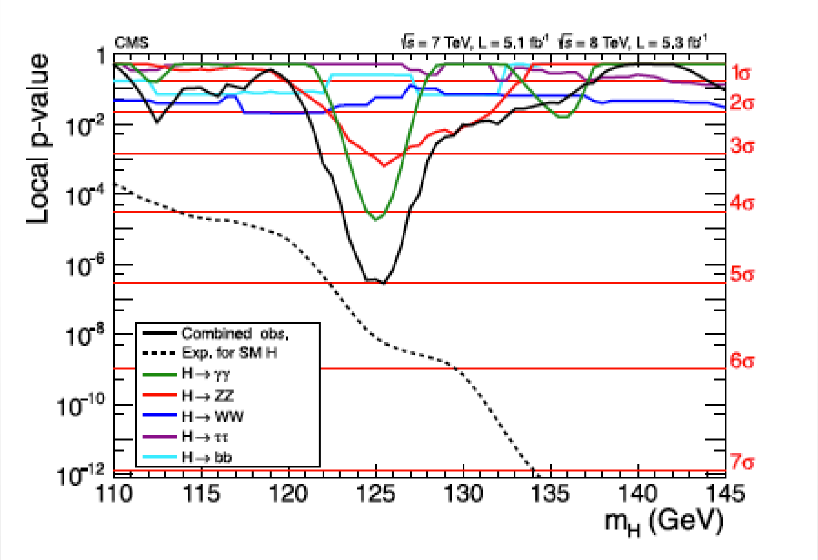


…to Atlas…



- ”Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC” (2012)

…and CMS



- “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC” (2012)

Modeling uncertainty: A place for Bayes?

Measurement

- A possible formulation of the problem: given data x_0 gathered via measuring procedure *M*, what range of values may be reasonably attributed to the measurand μ?
- Inference is based a on model of *M*.
- *M* and x_0 yield a model-based argument with conclusion $\mu = \mu_0$, where $\mu_0 - \epsilon_1 \leq \mu_0 \leq \mu_0 + \epsilon_2$.

Statistical uncertainty

- Variability of the outcomes of M (due to finite data) supports the use of frequentist methods to characterize the range of values (confidence intervals).
- The result is an estimate of the statistical uncertainty of the measurement result.
  - e.g., $\sigma_{t\bar{t}} = 187 \pm 11(\mathrm{stat.})\mathrm{pb}$
- Statistical uncertainty estimates are based on assumed knowledge about M – the distribution of the random variable.

Systematic Uncertainty

- Generally, the model of *M* itself is subject to some uncertainty:
  - input parameter values
  - detector resolution
  - simulation dependence
  - etc.
- Not all of these kinds of uncertainty lend themselves easily to analysis in terms of variability.
- This is not uncertainty due to the probability distribution but uncertainty *about* the probability distribution

Systematic uncertainty from robustness analysis

- To increase the cogency of a measurement argument:
  - Consider modifications of M that are compatible with background knowledge.
  - Determine what differences such modifications make to the range of values assigned to μ.
  - Report the range of such differences as the systematic uncertainty.
  - E.g, $\sigma_{t\bar{t}} = 187 \pm 11(\text{stat.})^{+18}_{-17}\,(\text{syst.})\text{pb}$
- Does background knowledge warrant attaching greater weight to some possible modifications than others?
- If so, this might be achieved by introducing a prior probability.

Problems for a Bayesian approach

- What kind of knowledge is required to warrant assigning a prior?
- Does this require giving up on frequentist concerns with keeping control over error rates?
- How should we interpret the resulting probabilistic statements?
- How should statistical and systematic uncertainties be combined?

Conclusions

- Statistical argumentation is central to the production of knowledge in HEP.
- Frequentist statistical tools provide useful means to meet the practical demands of argumentation in HEP.
- Significant unanswered questions remain about how to use probabilistic language to characterize uncertainty; frequentism and Bayesianism both seem to have something to offer, but is either adequate on its own?
- This final problem is necessarily interdisciplinary.
    - (Physicists working on systematic uncertainty in LHC experiments: I would like to talk to you!)

Thank you!