

# PROBATIVE FOUNDATIONS FOR BAYESIAN STATISTICS?

DAVID MWAKIMA

LOGIC AND PHILOSOPHY OF SCIENCE

UNIVERSITY OF CALIFORNIA

IRVINE, CA USA

OVERVIEW OF RESEARCH

## 1 Introduction

## 2 Background and Overview of Proposed Research

# Outline of my presentation

**1** Introduction

**2** Background and Overview of Proposed Research

## 1 Introduction

## 2 Background and Overview of Proposed Research

# What got me interested?

## The Centrality of Evidence

1. In epistemology and metaphysics, “A wise man proportions his belief to the **evidence.**” Hume (1748)
2. “Quality of Evidence” and the contemporary scientific realism debate.

*[Laudan's pessimistic meta-induction argument] disregards potentially important differences in **the quality and quantity of evidence** there is for current theories (differences that would justify treating current theories as more supported by **available evidence** than past theories were by the then **available evidence**); but also because it makes a mockery of looking for **evidence for scientific theories!***

*Psillos (2018)*

# What got me interested?

## The Centrality of Evidence

2. “Quality of Evidence” and the contemporary scientific realism debate.

*The realism debate itself is most fundamentally concerned with whether there is any general or categorical **variety of** empirical success or **evidential support** that serves as a **reliable indicator** that a theory (or its privileged parts) will be retained and ratified throughout the course of further inquiry.*

*Stanford (2021)*

## What Everyone has Been Talking About: Perrin's evidence for the Atomic Theory

*I've been suggesting all along, on the Second Philosopher's behalf, that **the evidence** involved in establishing the atomic hypothesis **wasn't just more of the same**, but **a new type of evidence** altogether, what we've been calling 'detection.'*

*Maddy (2007, 405)*

# Evidence and its Quality

1. Informal: epistemology, Quine, Maddy etc.
2. Formal: inductive logic and formal epistemology, Carnap etc.
3. Semi-formal: Achinstein, Roush
4. **Statistics: statistical evidence:** Bayes Factors, odds ratios, likelihood ratios, Mayo's Severity Function, p-values?



# Royall's Three Questions: Choosing a Paradigm to Assess the Quality of Evidence

1. What should I believe, now that I have this observation?  
**Bayesianism**
2. What should I do, now that I have this observation?  
**Frequentist/Classical Statistics**
3. How do I interpret this body of observation as evidence?  
**Likelihoodism**

# Likelihoodism

Is this all you need?

## 1. The Law of Likelihood

Evidence  $E$  for  $M_1$  is stronger than the evidence  $E$  for  $M_2$  if  $p(E|M_1) > p(E|M_2)$ , i.e., the likelihood ratio  $\frac{p(E|M_1)}{p(E|M_2)} > 1$

## 2. Likelihood Principle

- ▶ In making inferences or decisions about  $\theta$  after  $x$  is observed, all relevant experimental information is contained in the likelihood function for the observed  $x$ . Furthermore, two likelihood functions contain the same information about  $\theta$  if they are proportional to each other (as functions of  $\theta$ ).
- ▶ All the information necessary and sufficient for quantifying how data give rise to statistical evidence is contained in the likelihood functions.

# What more do we need?

## Mayo's Minimal Requirement for Severity

### 1. **Severe testing**

*A claim  $C$  is **severely tested** to the extent that it has been subjected to and passes a test that probably would have found flaws, were they present.*

### 2. **Error Probabilities**

*The capability of a method or rule for statistical inference to detect flaws, were they present is gauged using the methods **error probabilities**.*

## Minimal Requirement for Severity

If data  $x$  agree with a claim  $C$  but the method was practically incapable of finding flaws with  $C$  even if they exist, then  $x$  is poor evidence for  $C$ .

# Ending the “Statistical Wars”: Beyond Probabilism and Performance

1. **Probabilism:** A view of statistical inference and statistical evidence is probabilist if it relies on **probabilities** understood as degrees of actual or rational belief in a claim to arrive at absolute or comparative measures of the evidence one has for those claims. **Bayesianism and Likelihoodism**
2. **Performance:** A view of statistical inference and statistical evidence is **performance**-based if it appeals to long-run error rates of its procedures to justify the reliability of those procedures. **Neyman-Pearson Hypothesis Testing**
3. **Probativism:** A view of statistical inference and statistical evidence is probativist if it requires the control of long-run error rates and the **probing of methods**, tests and procedures for severity. **Mayo's Error Statistical Framework**

# Ending the “Statistical Wars”: Beyond Probabilism and Performance

## Against the Law of Likelihood

The Law of Likelihood does not satisfy the minimal requirement for severity.

- Recall the Law of Likelihood: Evidence  $E$  for  $M_1$  is stronger than the evidence  $E$  for  $M_2$  if  $p(E|M_1) > p(E|M_2)$ , i.e., the likelihood ratio  $\frac{p(E|M_1)}{p(E|M_2)} > 1$ .
- Call an alternative hypothesis  $H_1$  **Gellerized** just in case the probability of a likelihood ratio in favor of  $H_1$  over  $H_0$  is **maximal**.
- One can show that a Gellerized alternative hypothesis  $H_1$  always exists for any  $H_0$  (that is not itself a Gellerized hypothesis for the data  $\mathbf{x}$  at hand).

# Ending the “Statistical Wars”: Beyond Probabilism and Performance

## Against the Likelihood Principle

The Likelihood Principle doesn't satisfy the minimal requirement for severity.

- One consequence of the Likelihood Principle is “the irrelevance of optional stopping rules.”
- In an ideal case, a scientist ought to report the **actual** statistical significance of her results — the p-value — only based on her original sample size or fixed number of trials  $n$ .
- Does “trying and trying again” until one gets a statistically significant result constitute a questionable research practice? **Bayesians and Likelihoodists say, “It doesn't matter!”**
- “[D]ata do not speak for themselves.” (Mayo, 2018, 439)

# The Challenge for Bayesian Statistics

*For the error statistician, as long as an account is restricted to priors and likelihoods, it still **leaves out the essential ingredient for objectivity: the sampling distribution**, the basis for error probabilities and severity assessments. Classical Bayesians, both subjective and default, reject this appeal to “frequentist objectivity” as solely rooted in claims about long-run performance. **Failure** to craft a justification in terms of **probabativeness** means that there’s uncharted territory, waiting to be developed.*  
Mayo (2018, 231)

Can the challenge be met?

## This is My Proposed Research Project

1. **Which philosophy?** Subjective Bayesianism vs. Objective (Default/Pragmatic) Bayesianism and the Likelihood Principle.
2. Must we **unify** or can a subjective Bayesian statistician **accommodate** the probativist requirements for statistical evidence?  
*The idea of error statistical foundations for Bayesian tools is not as preposterous as it may seem. (Mayo, 2018, 28 – 29)*
3. How exactly does Mayo's framework, which requires **the sampling distribution for objectivity**, allow us to get beyond the statistics wars by being general enough to provide a conceptual or philosophical background for Bayesian statistics?



## References

- Hume, David (1748). *An Enquiry Concerning Human Understanding*. Oxford University Press. 1999 text edited by T. Beauchamp.
- Maddy, Penelope (2007). *Second Philosophy: A Naturalistic Method*. Oxford University Press.
- Mayo, Deborah G. (2018). *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press.
- Psillos, Stathis (2018). "Realism and Theory Change in Science". In *The Stanford Encyclopedia of Philosophy* (edited by Edward N. Zalta).
- Stanford, P. Kyle (2021). "Realism, Instrumentalism, Particularism: A Middle Path Forward in the Scientific Realism Debate". In *Contemporary Scientific Realism: Challenges from the History of Science* (edited by Timothy Lyons and Peter Vickers). Oxford University Press.

## Appendix: More on Severity

### Strong Requirement for Severity

If  $C$  passes a test that was highly capable of finding flaws or discrepancies from  $C$ , and yet none or few are found, then the passing result,  $x$ , is evidence for  $C$ .

- We have evidence for a claim  $C$  just to the extent it survives a stringent scrutiny.

### Minimal (Weak) Requirement for Severity

If data  $x$  agree with a claim  $C$  but the method was practically incapable of finding flaws with  $C$  even if they exist, then  $x$  is poor evidence for  $C$ .

- We do not have evidence for a claim  $C$  if there was no stringent scrutiny.

## Appendix: Severity Interpretation for Negative Results

- (a) **Low:** If there is very **low** probability that  $d(x_0)$  would have been larger than it is, even if  $H_1$  is true, then  $H_0$  passes with low severity:  $SEV(H_0)$  is low.
- Your test wasn't very capable of detecting discrepancy even if it existed. So when it is not detected, it is poor evidence of its absence.
- (b) **High:** If there is a very **high** probability that  $d(x_0)$  would have been larger than it is, were  $H_1$  is true, then  $H_0$  passes the test with high severity:  $SEV(H_0)$  is high.
- Your test was highly capable of detecting discrepancy if it existed. So when it is not detected, it is a good indication of its absence.

## Appendix: Severity Interpretation for Significant Results

- (a) **Low:** If there is a fairly **high** probability that  $d(x_0)$  would have been larger than it is, even if  $H_0$  is true, then  $d(x_0)$  is not a good indication for  $H_1$ :  $SEV(H_1)$  is low.
- (b) **High:** If there is a very **low** probability that so large a  $d(x_0)$  would have resulted, if  $H_0$  were true, then  $d(x_0)$  indicates  $H_1$ :  $SEV(H_1)$  is high.

# Appendix: Ending the “Statistical Wars”: Beyond Probabilism and Performance

## Against the Likelihood Principle

The Likelihood Principle doesn't satisfy the minimal requirement for severity.

- One consequence of the Likelihood Principle is “the irrelevance of optional stopping rules.”
- Let  $s$  denote the number of favorable outcomes (each with a probability 0.5) of a sequence of Bernoulli trials and  $n$  the number of trials.
- Continued on next slide....

# Appendix: Ending the “Statistical Wars”: Beyond Probabilism and Performance

## A Tale of Two Experiments

### Experiment 1

1. **Stopping Rule:** Stop experiment after 16 favorable outcomes.
2. Suppose the 16th favorable outcome occurs on the 24th Bernoulli trial.
3.  $P(n = 24) = \binom{24-1}{16-1} (0.5)^{16} (0.5)^8$
4. Attained significance level (p-value) is **0.077**.

### Experiment 2

1. **Stopping Rule:** Stop experiment after 24 Bernoulli trials.
2. Suppose you obtain 16 favorable outcomes.
3.  $P(s = 16) = \binom{24}{16} (0.5)^{16} (0.5)^8$
4. Attained significance level (p-value) is **0.032**.

## Appendix: Objective Bayesianism vs. Subjective Bayesianism And the Likelihood Principle

1. Objective (Default/Pragmatic Bayesianism): Priors are chosen such that the likelihoods based on the experiment/data dominate the computation of the posteriors. Box and Tiao (1973, 44)
2. The prior density is **locally** uniform over the parameter space.
3. **Jeffrey's prior**: A prior distribution for a single parameter  $\theta$  is approximately noninformative if it taken proportional to the square root of Fisher's information measure.
4. "The form of the prior **must** then depend on the expected likelihood." Box and Tiao (1973, 44)
5. Locally uniform prior for Binomial (fixed  $n$ ):  $\arcsin(\sqrt{\theta})$
6. Locally uniform prior for Negative Binomial (fixed  $s$ ):  
 $\log\left(\frac{1-\sqrt{1-\theta}}{1+\sqrt{1-\theta}}\right)$

## Appendix: Bayes Factor

Write equation (1) and (2) as:

$$\pi(M_1|D) = \frac{p(D|M_1)\pi(M_1)}{p(D)} \quad (1)$$

$$\pi(M_2|D) = \frac{p(D|M_2)\pi(M_2)}{p(D)} \quad (2)$$

$$\frac{\pi(M_1|D)}{\pi(M_1)} = \frac{p(D|M_1)}{p(D)} \quad (3)$$

$$\frac{\pi(M_2|D)}{\pi(M_2)} = \frac{p(D|M_2)}{p(D)} \quad (4)$$



## Appendix: Bayes Factor

$$\frac{\frac{\pi(M_1|D)}{\pi(M_1)}}{\frac{\pi(M_2|D)}{\pi(M_2)}} = \boxed{\frac{p(D|M_1)}{p(D|M_2)}} \quad (5)$$

The right hand side of equation (5) can be used to quantify the **relative predictive accuracy** of our models. This quotient is the **Bayes Factor**.

## Appendix: Bayes' Theorem Continuous Case

$$\pi(M_1|D) = \frac{\int_{\Theta} p(D|\theta)\pi(\theta|M_1)d\theta\pi(M_1)}{\sum_{i=1}^n \int_{\Theta} p(D|\theta)\pi(\theta|M_i)d\theta\pi(M_i)}$$

1.  $\Theta$  is the parameter space.<sup>1</sup>
2.  $\pi(\theta|M_i)$  is the prior on the parameter(s) given model  $M_i$ .
3.  $\pi(M_i)$  is the prior on model  $M_i$ .
4. Bayes Factor  $M_1$  vs.  $M_2$  is given by:

$$\frac{\int_{\Theta} p(D|\theta)\pi(\theta|M_1)d\theta}{\int_{\Theta} p(D|\theta)\pi(\theta|M_2)d\theta}$$

---

<sup>1</sup>See Handbook(2011, 595ff.)