

Final papers 6014
DUE May 11: 12-14 pages (1.5 spaced)

(Please come to my office to discuss your papers and/or write by email.)

The assignment. You are to imagine you are writing an article for a special issue of the *Journal of Comparative Statistical Methodology: An interdisciplinary journal on philosophy, biology, and economics*. The special issue is on “Current Controversies in Statistical Inference”. The (imaginary) CFP (call for papers) is described in detail below. It is *not* intended that all items listed under a theme in bullets be discussed. Many of them are different ways to get at the same theme. Nor is it assumed that you remain with a single topic (in bold). You are all good writers, and I trust you to mix and match, go broad or hone in on a few things. So long as your paper is on the “current controversies” topic of the special issue, and you provide the kind of philosophical analysis the journal editors are looking for (see 4th bullet). *You are all capable of a publishable paper on the topic, and that is why I’m giving this assignment.*

- Your final paper should include 1 new item from the “Captain’s Library” that you will read (at least in part). Once I know what you’re writing on, I can advise.
- You can use up to half the material from your short essays, but couched within the purpose of your final paper.
- Include Assignments 1-3 and your short essays when you submit your final papers (*electronically*).
- *Do a philosophical analysis:* You’re writing for a philosophical journal. What’s unique about giving a philosophical analysis of rival views and criticisms is the way it circles around opposing views and proceeds in stages to develop a strong argument. For example, you might put forward a position or criticism giving it a generous (or “charitable”) interpretation. You then raise criticisms or questions of that position, followed by a response by the advocate of the initial position. You go around again, and finally give your position and why (which might be that both sides are missing something).

Begin with an opening paragraph and summary explaining what you will discuss, what questions you will raise, or arguments you will consider. Why is your topic important for the mission of this special issue and what new illumination or perspectives will you provide?

Recommendations: You are all good writers, so proceed any way you prefer. You might wish to write a numbered outline so that your paper proceeds in logical order without going back and forth. Estimate how many pages for each section (There’s no firm limit; if quotes make it long, that’s fine.) Given the outline, start writing what you definitely want to say under each heading, and then place these items under their appropriate number. Your paper will be at least half written! I recommend you begin by rereading 6.7 Farewell Keepsake, and reviewing all the souvenirs from our reading (links are on the syllabus).

HAVE FUN!

CFP: April 2023

Call For Papers: Journal of Comparative Statistical Methodology: An interdisciplinary journal on philosophy, biology, and economics.

JCSM plans to run a special issue devoted to current controversies in statistical inference. We invite papers by graduate students and post docs who have studied the philosophical foundations of statistics and can illuminate our readers on the arguments and philosophical presuppositions underlying today's debates. The Journal is especially interested in articles that will be useful for early career researchers embarking on work in areas employing statistical methods, who need to understand today's controversies in statistics between and within frequentist error statistical, likelihoodist, and Bayesian methods. The goal of this special issue is to enable our readers to critically examine the different positions for themselves, avoiding groupthink and bandwagons. The Journal is also interested in articles that explore, in a general way, the differences in the overall conception of science held by Bayesians and frequentist error statisticians regarding the nature of learning from data, whether or not evidence should be comparative (likelihoodists and Bayesians) or not (statistical significance testers), objectivity/subjectivity, and the relevance/irrelevance of error probabilities.

Replication crisis and paradoxes: Controversies over principles of inference

The replication crisis has put the spotlight on statistical methodology. One goal for contributions to this special issue is to shed light on how the replication crisis connects to any of the following issues: redefining significance, likelihood principle, Fisher/Bayes disagreement, default Bayesian priors, spike and smear priors; P-values vs. posterior probabilities.

- Examples of proposed remedies for lack of replication are preregistered reports, replication, and open science. Why is it supposed they would help with issues of data-dredging and multiple testing? Others claim that the replication crisis shows the need to change from P-values to Bayes factors. In what way are these two routes to improving replication in tension? (A replication paradox.)
- What is the main gist of the likelihood principle? Why does accepting or rejecting this principle lead to disagreement as to whether multiple testing and optional stopping are relevant to evidence? Is this a matter of a difference in one's philosophy of inference/evidence?
- Why does appealing to error probabilities result in violating the likelihood principle? Is concern about optional stopping tantamount to making private intentions relevant to evidence?
- We are interested in the author's view of the debate about the likelihood principle, either in its relation to replication or in its own right?

Contrasts between and within Bayesian and frequentist error statistical approaches.

Bayesians assign probabilities to fixed parameters and hypotheses. Since parameters are not random, probability assignments to them are not frequentist (except in empirical Bayes, where the parameter is viewed as coming from a superpopulation with known distribution). The question therefore arises: How should these Bayesian probabilities be interpreted, obtained and used?

- A well-known Bayesian (Gelman) declares in a joint paper that “most of the standard philosophy of Bayes is wrong” (p. 10, n. 2). Papers that explain the gollimaufry of meanings given to prior probabilities (and how to obtain them) are much needed.
- Can distinguishing probabilism, performance, and probativism enable holding distinct methods for different contexts? Or not?
- How would you contrast the assessment of statistical model assumptions in frequentist versus Bayesian accounts? For some contemporary Bayesians (“falsificationist” Bayesians like Gelman, Mayo 2018, Section 6.6), the prior is part of the model and thus is no different from the ordinary frequentist model assumptions. How would you critically assess this? (SIST 6.6)
- Some are entering fields that use Bayesian statistics but they are unsure whether they need to satisfy principles such as *weak-repeated sampling*, *likelihood principle*, *stopping rule principle*. Clinical trialists (Ryan et al 2020) are conflicted between agency requirements (in the case of adaptive designs) and the Bayesian philosophy. The Journal is interested in articles explaining these.
- Bayesians and likelihoodists ‘condition’ on the data, once it is available, whereas error statisticians consider outcomes other than the one observed in using error probabilities and tail areas. As a result, it is said (e.g., Kadane) that Bayesians and error statisticians flip the role of probability (in terms of what is random and what is fixed). What is this flipping about?

Controversies regarding frequentist error-statistical methods. Frequentist error statisticians use probability to assess and control a method’s ability to avoid erroneous interpretations of data in repeated use.

- What are fallacies of rejection and acceptance? The large n problem? How does the severity interpretation propose to tackle the main criticisms of P-values and of N-P accept/reject tests (Mayo 2018)? What is your view?
- A major criticism of Fisherian (simple) statistical significance tests is that failing to find a low P-value is claimed to be uninformative: one just says, “no statistical significance was found”. What is your view of using power analysis to avoid fallacies of nonsignificance (J. Cohen’s work)? Is it too coarse (Mayo)?
- Confidence intervals and severity are others ways to set upper bounds in the case of nonsignificant results. Discuss.

Epistemological value of error probabilities

- Why do some frequentist error statisticians (e.g., Mayo) criticize the standard view of null hypothesis significance testing (NHST), and reject the existing formulation of N-P tests as providing tools with good long-run error control?
- What are the main differences between Fisherian (simple) statistical significance tests and N-P tests (and confidence intervals)? Why do some claim today’s frequentist practice is an inconsistent hybrid between the two (Gigerenzer)? How does severity propose to get beyond the “inconsistent hybrid” charge?
- A severity assessment uses error probabilities to qualify particular inferences by considering a method’s capacity to probe errors. How is this intended to give an inferential assessment? While akin to a Popperian degree of corroboration, this is at odds with using probability to quantify degree of (actual or rational) belief or support to

statistical hypotheses (probabilism). Explain the difference. Can error probabilities still be inferential or only for performance)?

Redefining statistical significance and its controversies

- What is the controversy about whether P-values “exaggerate evidence”? (Berger and Sellke 1987; Casella and Berger 1987). If a P-value of .05 exaggerates evidence, then doesn't requiring $P = .005$ result in underestimating the effect? Explain.
- What is the call to “redefine” significance? Is it merely recommending a lower P-value, or advocating using Bayes factors? Many say it is both. Can you explain?
- The recommendation to lower P to .005 was popularized by V. Johnson's Bayes factor approach; what is the argument that this actually exaggerates the evidence by allowing inferences to alternatives with only .5 severity (Mayo 2018, 262-3)?

Posterior predictive value and the “diagnostic screening model” of statistical tests

- Ioannidis' (2005) famous article argued that “most scientific findings are false” based on an assumed high “base rate” of true null hypotheses. Many use Ioannidis' argument to pinpoint the blame for lack of replication. Explain and critically evaluate.
- How is the (Bayesian) computation in the “diagnostic screening model” related to the call to “redefine” statistical significance?
- Explain how the appeal to “base-rates” arises as a general criticism of frequentist error probabilities, including the post-data severity evaluation. (Howson 1997, Mayo 1997) Your view?

The 2016 policy statement on P-values and replication:

Our Journal would be interested in a paper that discussed and critically analyzed each of the 6 principles in the 2016 statement (connecting to relevant points listed throughout this CFP.)

- The Wasserstein et al. 2019 editorial, an opening to a special *American Statistician* issue on “a world beyond $P < .05$,” (first few pages) went much further than the 2016 policy statement and revised several of the 2016 principles. There has been a lot of confusion between the two documents (and a third ASA Task Force document (Benjamini et al. 2022) that our readers might be interested in).
-

An explanation and critical analysis of a leading recent article (touching upon key themes in this CFP): McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). “Abandon statistical significance.” *American Statistician*, 73, 235–245.