# Understanding Non-Bayesians: Sims (2010)

Francis J. DiTraglia

University of Oxford

Summer of Bayes 2024

# Goals of this Presentation[1]

1. Summarize the main points of Sims (2010) - Understanding Non-Bayesians
2. Relate them to the broader discussion of Bayesian vs. Frequentist inference.

[1] Unless otherwise indicated, all quotes are from Sims (2010).

# Fun Facts

- "Understanding Non-Bayesians" was originally written for the Handbook of Bayesian Econometrics in 2010.
- Oxford University Press objected to Sims posting a pre-print on his website.
- Sims favors open access and withdrew from the handbook in protest; the paper remains unpublished.
- In 2011 Sims was awarded the Economics Nobel for "understanding cause and effect in the macroeconomy."
- Take that OUP!

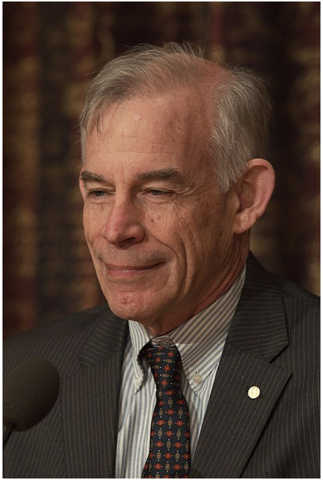

Figure 1: Chris Sims in 2011

# Motivation: Why isn't everyone Bayesian?

> *Once one becomes used to thinking about inference from a Bayesian perspective, it becomes difficult to understand why many econometricians are uncomfortable with that way of thinking. But some very good econometricians are either firmly non-Bayesian or (more commonly these days) think of Bayesian approaches as a "tool" which might sometimes be appropriate, sometimes not.*

Figure 2: Laplace in 1775

## Some Pedantry

Maybe we should call it "Bayes-Laplace" inference: Laplace developed what we would recognize as "Bayesian" inference.

# What is this paper about?

- Purpose: Articulate counterarguments to Bayesian perspective
- Some counterarguments are "easily dismissed"
- Others relate to deep questions about inference in infinite-dimensional spaces

*My conclusion is that the Bayesian perspective is indeed universally applicable, but that "non-parametric" inference is hard, in ways about which both Bayesians and non-Bayesians are sometimes careless.*

# Bayesian versus Frequentist Approaches

## Frequentist

- "Insists on a sharp distinction between unobserved, but non-random 'parameters' and observable, random data."
- "Works entirely with the probability distributions of data, conditional on unknown parameters–estimators and test statistics, for example–and makes assertions about the distribution of those function of the data, conditional on parameters"

## Bayesian

- "Treats everything as random before it is observed, and everything observed as, once observed, no longer random."
- "Aims at assisting in constructing probability statements about anything as yet unobserved (including 'parameters') conditional on the observed data."

# Let's unpack this a bit: Frequentists

- Condition on parameters; make probability statements that average over different *datasets* you could potentially observe.
- E.g. $X_1 ..., X_{100}$ is a random sample from a $\mathrm{N}(\mu, \sigma^2)$ population with $\sigma = 1$ known.
- "$\bar{X}_n \pm 1.96 \times \sigma/\sqrt{n}$ is a 95% confidence interval for $\mu$."
- Translation: In 95% of the datasets we could possibly observe, the sample mean will land within about $\pm 0.2$ of the true (fixed and unknown) value of $\mu$.
- The observed interval $\bar{x} \pm 0.2$ either contains $\mu$ or doesn't: nothing is random after we have observed the data.
- Traditional (Neyman-Pearson) inference is *pre-experimental*: inductive behavior rather than inductive inference.

# Let's unpack this a bit: Bayesians

- Condition on *observed data*; make probability statements that average over different *parameter values* that could potentially have generated the data.
- E.g. $X_1 ..., X_{100}$ is a random sample from a $\mathrm{N}(\mu, \sigma^2)$ population with $\sigma = 1$ known.
- Need a prior: just for simplicity choose a "vague" one e.g. $\mu \sim \mathrm{N}(0, 10000)$
- "The posterior distribution of $\mu$ is (approximately) $\mathrm{N}(\bar{x}, 1/100)$, so the 95% highest posterior density interval (HPDI) for $\mu$ is approximately $\bar{x} \pm 0.2$."
- Translation: Given the observed data, there is around a 95% probability that the population mean $\mu$ lies within $\pm 0.2$ of the sample mean $\bar{x}$.
- The observed sample mean $\bar{x}$ is fixed and known; the population mean $\mu$ is unknown and treated as random.
- Bayesian inference is *post-experimental* (conditional): inductive inference under an assumed model given observed information.

# Differing Interpretations of Probability

- Crucial background to the differences between Bayesians and Frequentists.
- Math is the same either way (Kolmogorov Axioms) but meaning is different.
- Bayesians: **"Belief-type"** (aka Epistemic) interpretation.
- Frequentists: **"Frequency-type"** (aka Aleatory) interpretation.
- Sims doesn't discuss this so the next two slides (including quotes) follow Chapter 11 of An Introduction to Probability and Inductive Logic by Ian Hacking.

# Differing Interpretations of Probability[2]

## Frequency-Type (Frequentists)

- What is the probability that a coin flip will come up heads?
- "We seem to be making a completely factual statement about a material object, namely the coin (and the device)
- In principle, we could repeat the experiment many times under identical conditions.
- This use of the term probability "is related to ideas such as frequency, disposition, tendency, symmetry, propensity."
- Some people call this "objective" probability but that's a loaded term: don't let words do your thinking for you!

## Belief-Type (Bayesians)

- What is the probability that Joe Biden will win the 2024 election?"
- One-time event: doesn't make sense to talk about frequency / propensity.
- Some people call this "subjective" probability but that's also a loaded term!

[2]Quotes on this slide are from Chapter 11 of Hacking's An Introduction to Probability and Inductive

# Learn some $*@%&!# physics before you talk to me about coin flips![3]

[3] See Chapter of 10 of Jaynes' Probability Theory: The Logic of Science for more discussion.

# Why does Sims put the word “parameters” in quotes?

Poirier (1996) textbook on econometrics: chapter 5?

# Implications for Decision-Making

- Bayesian inference feeds naturally into decision-making under uncertainty
- Frequentist analysis does not directly provide probabilities for decision-makers

# Easily Dismissed Objections

1. **"Bayesian inference is subjective"**
   - Bayesians can take an "objective" approach by describing the likelihood
   - Good frequentist practice also involves informal use of prior beliefs
2. **"Bayesian inference is harder"**
   - Often easier to characterize optimal small-sample inference from Bayesian perspective
   - Frequentist asymptotic results can often be given Bayesian interpretations

# Less Easily Dismissed Objections

### Handy methods that seem un-Bayesian

- IV, GMM, sandwich estimators, kernel methods
- Can be given limited information Bayesian interpretations
- Involve implicit Bayesian judgments in asymptotic theory

# Challenges in Non-parametrics

- Infinite-dimensional parameter spaces
- Consistency issues in Bayesian inference
- Pitfalls in high-dimensional spaces:
  - Priors can be unintentionally dogmatic
  - Importance of careful prior specification

# Example: Angrist and Krueger (1991) Quarter of Birth

- The Wasserman problem is about non-parametrics and you can read about it here.
- But we don't need anything too exotic to see the issues Sims is talking about.
- If you're not an economist and don't know what instrumental variables is, here's a very quick introduction.
- Give the introduction.
- Then make the point of Chamberlain & Imbens (1996)
- Point out that the Frequentist solution is also terrible in this case since it corresponds to an insane prior!
- Useful dialogue between Bayesians and Frequentists: what prior does the frequentist solution correspond to? Frequency properties of Bayesian estimators?

# Example 1: The Wasserman Problem

- Setup: Observing $(\xi, R, Y)$ with unobserved $\theta$
- Goal: Estimate $\psi = \mathbb{E}[\theta]$
- Bayesian approaches:
  1. Independence case
  2. Dependence case (sieve method)
  3. Limited information approach

# Critique of Wasserman's Conclusions

I probably still want to mention these points, but I don't really want to get into the Wasserman example since it won't be as familiar to the audience.

- Bayesian methods are not necessarily insensitive to data
- Importance of appropriate prior specification
- Pitfalls of high-dimensional parameter spaces

# Example 2: Robust Variance Estimates in Regression

Not sure how much I should say about this one, but if I do mention it then it might be worth mentioning the Leamer "White-washing" stuff along with the paper where he talks about the "sandwich" estimator versus GLS and something about when the point estimates will change.

- OLS with sandwich covariance matrix
- Efficiency bounds (Chamberlain, 1987)
- When is OLS with sandwich appropriate?
  - Large samples
  - Likely nonlinear regression function
  - Interest in best linear predictor

# Conclusion

- Bayesian perspective is universally applicable
- Importance of careful modeling in high-dimensional spaces
- Pragmatic Bayesian approach:
  - Recognize limitations of asymptotic approximations
  - Consider model improvements when appropriate
  - Use OLS with sandwich judiciously

# Questions?