Introductory Statistics Curriculum

At The Crossroads of Bayesian and Frequentist

Ian McConachie

March 2023

Abstract

This report compares the two primary approaches to statistical reasoning in terms of their application to science and research. It provides justification for the inclusion of Bayesian ideas in curriculum and gives suggestions for how such ideas may be introduced.

Executive Summary

Introduction

The field of statistics has two main approaches: the Bayesian approach and the frequentist approach. This report investigates the current state of introductory, undergraduate-level statistics curriculum with relation to these approaches.

Bayesian vs. Frequentist

The main differences between the approaches can be summarized as:

- Bayesian inference allows incorporating prior beliefs, the frequentist approach does not
- Bayesian models are more computationally expensive than their frequentist counterparts
- Frequentist models must be changed depending on experiment design, Bayesian models are more easily adapted

Why Teach Bayesian?

Currently frequentist statistics dominant introductory curriculum, but there are several reasons why the Bayesian approach should get more attention which include:

- The increased prominence of Bayesian inference in academia
- The trend of manipulating and misusing frequentist methodologies leading to setbacks in the progress of science
- Bayesian models outperform their frequentist counterparts in a variety of common applications of statistics

Application to Curriculum

Given the importance of Bayesian inference in the field of statistics, introductory courses should adapt to include Bayesian methods in their content. It has been shown in several empirical studies that teaching Bayesian from students' first foray into the field can have positive learning outcomes. These courses may either be completely Bayesian or mix ideas from both approaches.

Conclusion and Suggestions

It is recommended that undergraduate courses that introduce statistics be modified so that they cover both the frequentist and Bayesian approach to inference. Doing this should increase statistical reasoning skills among the general public and decrease the prominence of statistical malpractice.

Contents

Introduction

This report seeks to investigate the possibility of teaching Bayesian statistics to undergraduate-level students in their first statistics classes. In doing this, we will examine several aspects of the two approaches and compare and contrast them through the lens of undergraduate education.

Intended Audience

This report is written for mathematics educators and administrators at institutions that offer undergraduate-level statistics courses. As such, a certain level of familiarity with statistics and mathematics as a whole is assumed in the report below.

Scope

This report covers the differences between Bayesian and frequentist statistics from a practical point of view, but does not give a mathematically robust definition of each approach. Instead it focuses on comparing the two in their application to science because this is how most students taking introductory statistics courses will use what they learn.

Background

In university math departments, introductory statistics is often among the most popular courses because of its application in a wide variety of fields outside of mathematics.The curriculum taught in these classes is extremely important as a misunderstanding of the basics of statistics continues to be a problem in many fields of research and industry.[21, 17] As machine learning and other forms of statistical reasoning become more popular, this shaky understanding of statistics will prove to be an even greater challenge in the near future.

Introductory undergraduate-level statistics courses have traditionally been taught using the frequentist approach to statistical reasoning. The other main approach to statistics—the Bayesian approach—-has historically been unfeasible for widespread use because of computational limitations. However, with the ubiquity of computing in the modern world Bayesian models are becoming more popular in research and industry.[4,11]

Brief History

Statistics as a discipline traces it origins back to the first formulations of probability theory by Gerolamo Cardano in the 16th and 17th centuries.[15]

Despite currently being the less popular, Bayesian statistics has a longer history than the frequentist approach. It is built upon Bayes Theorem (where it gets its name) which was first formulated by Thomas Bayes in the 1700s.[1] The French Mathematician Pierre-Simon

Laplace is credited with formalizing many of the statistical methodologies originally based on Bayes' work.[17]

In the early 1920s two mathematicians, Ronald Fisher and Jerzy Neyman, developed an alternative approach to probability and statistical reasoning which we now know as frequentist reasoning (also ironically sometimes referred to as ”classical” statistics).

Today the use of Bayesian statistics is increasing once again thanks to widespread computational power and the development of feasible random sampling algorithms such as Markov Chains.

Bayesian vs. Frequentist

The discipline of statistics is built on mathematical principles of probability and so the core difference between these two approaches ultimately comes down to their different interpretations of probability.

In the frequentist school of thought probability of an outcome is thought to be the frequency at which we get that outcome when we repeat the experiment many times. On the other hand Bayesian statisticians interpret the probability of an outcome as our belief that the outcome will occur.[12]

These definitions may seem very similar, but the small differences lead to major structural differences as a statistical framework is built upon them. Neither definition is right or wrong, but they have different implications for the mathematical side of statistics. Some of the most important distinctions between frequentist and Bayesian statistics are summarized below (Figure 1):

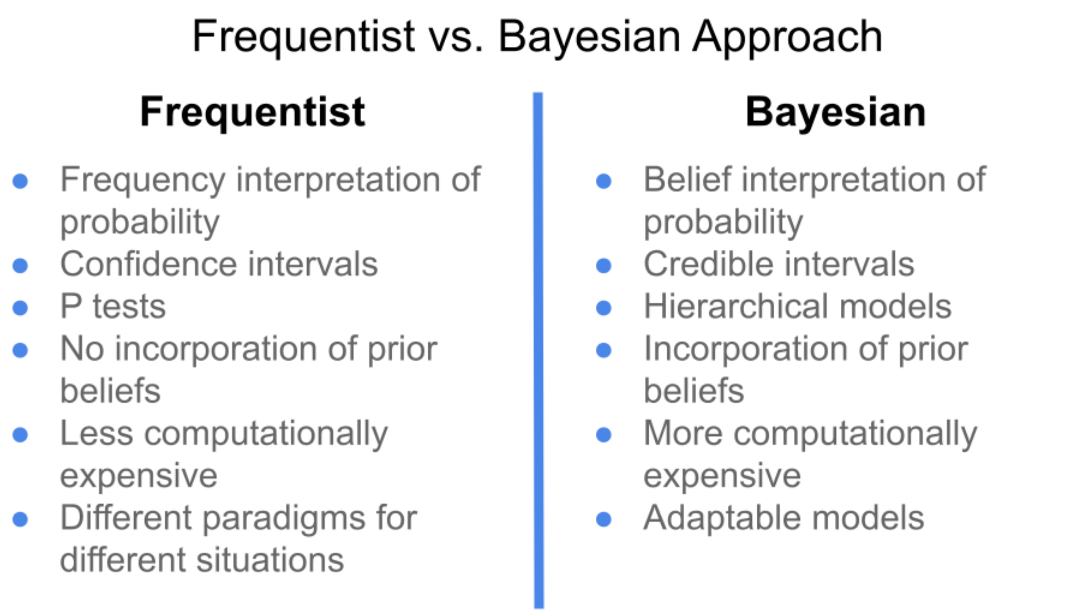

Figure 1: An overview of the differences between these two approaches.

In the sections below we will highlight some of these differences that will be important to us moving forward.

Incorporation of Prior Beliefs

One of the main reasons that the general scientific community moved from the Bayesian approach to the frequentist in the early 20th century was the argument that frequentist statistics were more "objective." This idea originated because Bayesian statistics allows statisticians to incorporate prior beliefs about parameters into models, which some see as going against the ethos of science.

These prior beliefs are expressed as prior distributions have an affect on the ultimate probability distribution (called the posterior distribution in Bayesian circles). Oftentimes prior knowledge can give us a much better idea of the likelihood of an event occurring—for instance a diagnostic test (even if very accurate) is much more likely to yield a false positive if the disease it is testing for is rare.[11] In Bayesian inference, we may include the likelihood of the disease showing up in the general population as a prior distribution, while it is much more difficult to build this idea into a frequentist model.

Computational Cost

Due to the use of prior distributions, it is often impossible to analytically find the true posterior distribution for a Bayesian model. This means that in practice statisticians often use random sampling from the prior to approximate the posterior. This technique is effective, but comes at a cost in terms of computation.

Random sampling from the prior is often done using Markov Chain Monte Carlo (MCMC) simulations which are conducted using sophisticated algorithms. MCMC simulations take many, many computations to be effective, so they draw on more computational power than simply examining the distribution of observed data as is done in frequentist inference.[6]

Model Adaptability

Another key difference between the two approaches is that frequentist statistics is built upon a specific part of probability theory called sampling theory.[22] This means the process used to collect data must be constant and repeatable.

This has ramifications in the practical application of statistics because frequentist analysis is not possible unless you are aware of the experiment design. When the same data is collected, but with different methods of sampling frequentist inference can lead to significantly different conclusions.[11]

On the other hand, Bayesian statistics allows for inference to be made on data collected without necessarily knowing *how* the data was collected. This is becoming increasingly useful, as data that is not collected using a well-defined experimental design is something we see more and more in the modern information age.

Why Teach Bayesian?

Increased Prominence of Bayesian Methods

One of the primary reasons statistics courses are some of the most popular math classes in undergraduate programs is that statistics is often necessarily for quantitative analysis which appears in research in almost every academic field.

In the past, when Bayesian statistics was relatively uncommon in research communities, it made sense to only teach frequentist statistics to undergraduates because that is the kind of reasoning they would see in any academic paper they read. However as the methods behind Bayesian models have improved, more and more researchers are opting to use Bayesian inference in empirical studies.

The figure below (Figure 2) shows the change in a proxy for the use of Bayesian statistics in academia from 1974 to 2022.

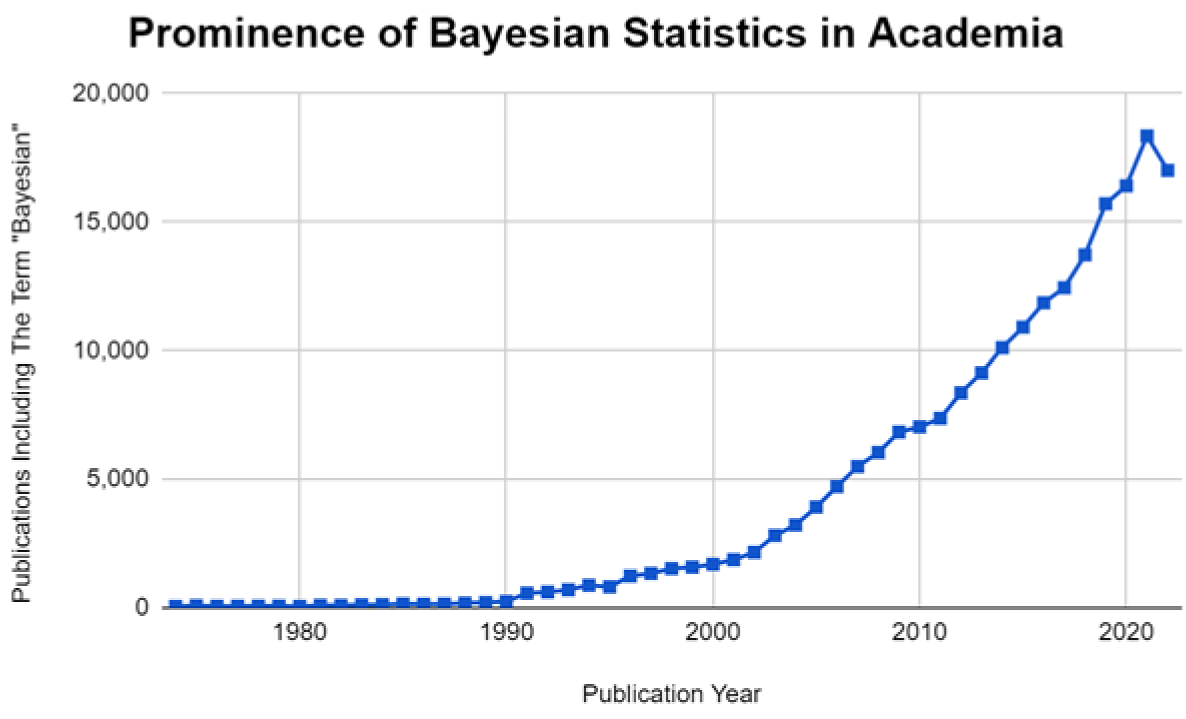

Figure 2: Line graph showing increase in number of papers that reference the term "Bayesian." Data collected from the Web of Science database compiled by Clarivate.[8]

The Problem With The Frequentist Approach

It is no secret that the scientific research community has a problem with the misuse and abuse of statistical inference. A quick google search will yield results from well respected journals like Nature, to popular science journalism sites like Scientific American.[7,2]

It is naïve to believe that using the Bayesian statistics would solve all of our problems with statistics—especially when there exists the purposeful misuse of statistics for personal gain. However, there are problems that arise more in frequentist statistics which could be partially addressed by reevaluating our statistics curriculum at the undergraduate level.

P Values

One of the most common applications of frequentist inference comes in the form of the P-test for statistical significance. Reading any scientific journal in any field you will very likely run into a statement that says "with significance $p < 0.05$ we reject the null hypothesis and conclude that..." It is a phrase that is drilled into young scientists' minds from their very first introduction to the practice of statistics. However, p-values and their place in scientific reasoning has come under increasing scrutiny as the field of statistics has evolved.

For starters, it should be noted that the de facto standard p-value of 0.05 that is seen in so many statistics textbooks and academic publications is completely arbitrary.

The origin of this metric for the binary classification of statistical significance can be traced back to a book from Ronald Fisher, one of the fathers of the frequentist approach (See Brief History Section).[9] In this book he posits 0.05 as a "convenient" delineation, but does suggest that it should be a standard.[9] In many ways the arbitrary selection of a p-value to consider significant takes away from arguments that frequentist inference is a more objective approach than Bayesian inference.

Further problems arise when the practice of P-tests is applied by researchers. Among most research communities, there are incentives to publish and most journals systematically favor articles that show positive (statistically significant) results—a phenomenon called "publication bias".[16] This has led to the widespread practice of "p-hacking."

"P-hacking" is a popular term used to describe a family of statistical abuses that take advantage of the weaknesses of P-testing as a tool for inference. Although not always intentional, p-hacking is an extremely common in many fields. One study from the National Institute of Health found through a meta-analysis of thousands of openly available publications that p-hacking was a significant barrier to science in many fields—particularly in medical and psychological research.[13] (Ironically they concluded significance based on one-sided P-tests).

Statistical Design

Another issue we see in part because of the prominence of frequentist statistics is errors in statistical design. As we mentioned above (see Model Adaptability), frequentist inference requires statisticians take into account the architecture of the experiment when constructing tests for statistical inference.

This leads to problems because a lack of planning before collecting data can lead to irreversible errors and even completely invalidate a publication's findings.[21] Bayesian models are more flexible in the type of data they can use to make inferences which makes them less susceptible to these kind of errors.

In addition to the design of tests on the data, researchers must also choose metrics for significance and justify their choice *before* conducting the experiment. Certain features of frequentist inference make transparency about the timing of analytical decisions crucial and this is information that is easily omitted or intentionally fudged in publications.[21] In Bayesian models, much less decisions must be made a priori which leaves less room for mistakes and malpractice.

Comparing Performance

Although the frequentist approach became dominant in the world of research after its formalization in the early 20th century, Bayesian models were still used for practical problems in areas such as espionage and military tactical decisions.[20,8] The Bayesian approach to probability survived because it was often found to be more effective than other methods in real-world applications.

Of course, it would be a gross over-simplification to say that Bayesian models always outperform their frequentist counterparts. However, it is generally accepted that Bayesian inference is optimal if the assumptions it is built on hold and there is sufficient computational resources to approximate a posterior distribution. The latter consideration has been largely satisfied thanks to the ubiquity of very powerful computers. There exists more nuance when looking at the former.

Although the assumptions behind the practical application of Bayes theorem are not always satisfied, there have been many empirical studies which have demonstrated that in many real-world use-cases of statistics we can safely apply Bayesian inference. In particular studies have shown that Bayesian models perform better or equally well when compared to the frequentist approach in common experimental designs such as multiple treatment comparisons, longitudinal studies, and predictive regression.[10,5,18]

Application to Curriculum

The information above highlights the difference between the Bayesian and frequentist approach and details why Bayesian statistics is worth learning.

Introductory statistics courses at the undergraduate level are the only statistics courses that many students' will take in their educational careers—if not the only stats class they will take, these classes often serve as the foundation for any further understanding of the field. It then logically follows that Bayesian ideas should be at the very least introduced in such courses.

Feasibility

For the past century Bayesian statistics has been reserved for only higher-level stats students, so it is reasonable to ask if the ideas behind the Bayesian approach are too advanced to be introduced in students' very first courses on the subject.

At its core, the mathematics behind frequentist and Bayesian inference do not vary much in terms of difficulty. This is a subjective evaluation, but it is one that has been backed up by studies in which introductory stats classes were taught using the Bayesian approach with successful learning outcomes.[3]

Content of Courses

Recommending an overhaul of all statistics courses to completely switch from frequentist to Bayesian statistics is not only impractical, but also unnecessary.

Although Bayesian statistics is growing in popularity, it is still seen less often in research than the classical approach. We then would want to modify our course curriculum to not replace frequentist methodologies, but rather create new courses that introduce students to both simultaneously.

Many statisticians agree that both approaches have their place in the world of quantitative reasoning and ideally we can move forward with some combination of methodologies.[19]

Some may argue that this is too much content for a single introductory course and that is something to consider. However, based on the world's increased reliance on statistical reasoning and how often we see the discipline being misused—an extended introduction at the undergraduate level may be necessary.

Conclusion

The information presented in this report offers a glimpse at the current landscape of statistics. We have our two primary approaches to interpreting probability, Bayesian and frequentist, which have considerable impacts on the methodologies used in statistical reasoning. As it stands, the frequentist school of thought is favored in research and undergraduate classes, but trends are changing.

As we continue to see the Bayesian approach increase in popularity in a variety of applications, we must reexamine the comparison between models from each side. We see that frequentist models have been shown to be susceptible to statistical malpractice and misinterpretation through the use of p-hacking and incorrect test design. We also see that, even when both methods are used correctly, Bayesian models often match or outperform their frequentist counterparts.

Introductory statistics classes at the undergraduate level are key in many people's understanding of statistics. It is therefore recommended that Bayesian inference be included in the curriculum for these courses in some way.

The existence of established Bayesian pedagogical resources and demonstrated success of introductory Bayesian courses tells us that Bayesian ideas can be effectively incorporated into undergraduate curriculum. Bayesian inference should not completely replace traditional statistics teaching, rather both should be introduced side-by-side as much as that is possible.

Our world is increasingly influenced by statistics which makes it very important that each and every one of us has an understanding of the mathematics behind the methods. The current introductory statistics curriculum has fallen short in preparing students for real-world applications in research and industry.

By considering the Bayesian approach as a starting place we can re-frame undergraduate statistics curriculum and potentially make a change that will result in a large positive impact in both the world of statistics and humanity at large.

References

1. Bellhouse. (2004). The Reverend Thomas Bayes, FRS: A Biography to Celebrate the Tercentenary of His Birth. Statistical Science, 19(1), 3–32. https://doi.org/10.1214/088342304000000189

2. Bischoff, M. (2022, September 30). Statistics are being abused, but mathematicians are fighting back. Retrieved March 01, 2023, from https://www.scientificamerican.com/article/statistics-are-being-abused-but-mathematicians-are-fighting-back/

3. Bolstad, W. M. (2002). Teaching Bayesian Statistics To Undergraduates: Who, What, Where, When, Why, and How. Retrieved March 01, 2023, from https://www.stat.auckland.ac.nz/ iase/publications/1/3f2_bols.pdf

4. Boulanger, B., PhD, Carlin, B. P., PhD. (2021, December 21). How and why Bayesian statistics are Revolutionizing Pharmaceutical Decision making. Retrieved March 01, 2023, from https://acrpnet.org/2021/12/21/how-and-why-bayesian-statistics-are-revolutionizing-pharmaceutical-decision-making/

5. Carlin. (2013). Case study comparing bayesian and frequentist approaches for multiple treatment comparisons. Agency for Healthcare Research and Quality US.

6. Columbia University. (2023, March 13). Markov chain Monte Carlo. Retrieved March 01, 2023, from https://www.publichealth.columbia.edu/research/population-health-methods/markov-chain-monte-carlo

7. Conroy, G. (2020, September 08). How to spot dubious claims in scientific papers. Retrieved March 19, 2023, from https://www.nature.com/nature-index/news-blog/how-to-spot-dubious-claims-in-scientific-research-academic-papers

8. Dale. (2012). The Theory That Would Not Die–A Book Review. Notices of the American Mathematical Society, 59(5), 1–. https://doi.org/10.1090/noti839

9. Denworth, L. (2019, October 01). The significant problem of P values. Retrieved March 01, 2023, from https://www.scientificamerican.com/article/the-significant-problem-of-p-values/

10. Doll, Lauer, T. E. (2014). Comparing Bayesian and frequentist methods of fisheries models: Hierarchical catch curves. Journal of Great Lakes Research, 40, 41–48. https://doi.org/10.1016/j.jglr.2014.07.006

11. Hackenberger. (2019). Bayes or not Bayes, is this the question? Croatian Medical Journal, 60(1), 50–52. https://doi.org/10.3325/cmj.2019.60.50

12. Haslwanter, T., PhD. (2015, April). Bayesian vs. frequentist interpretation. Retrieved March 19, 2023, from http://work.thaslwanter.at/Stats/html/statsBayes.html

13. Head, Holman, L., Lanfear, R., Kahn, A. T., Jennions, M. D. (2015). The extent and consequences of p-hacking in science. PLoS Biology, 13(3), e1002106–e1002106. https://doi.org/10.1371/journal.pbio.1002106

14. Lai, M. (2019, December 13). Course handouts for Bayesian Data Analysis class - Ch. 1. Retrieved March 01, 2023, from https://bookdown.org/marklhc/notes_bookdown /introduction.html

15. Leung, M., PhD (2021). Probability Research Information and Statistics Education Modules - The Beginning. Retrieved March 01, 2023, from https://www.math.utep.edu/ faculty/mleung/probabilityandstatistics/beg.html

16. Nair. (2019). Publication bias - Importance of studies with negative results. Indian Journal of Anaesthesia, 63(6), 505–507. https://doi.org/10.4103/ija.IJA_142_19

17. Nuijten, Hartgerink, C. H. ., van Assen, M. A. L. ., Epskamp, S., Wicherts, J. . (2016). The prevalence of statistical reporting errors in psychology (1985–2013). Behavior Research Methods, 48(4), 1205–1226. https://doi.org/10.3758/s13428-015-0664-2

18. Pérez-Millan, Contador, J., Tudela, R., Niñerola-Baizán, A., Setoain, X., Lladó, A., Sánchez-Valle, R., Sala-Llonch, R. (2022). Evaluating the performance of Bayesian and frequentist approaches for longitudinal modeling: application to Alzheimer's disease. Scientific Reports, 12(1), 14448–14448. https://doi.org/10.1038/s41598-022-18129-4

19. Steinhardt, J. (2012, October 31). Beyond bayesians and frequentists. Retrieved March 01, 2023, from https://jsteinhardt.stat.berkeley.edu/files/stats-essay.pdf

20. Simpson. (2010). Edward Simpson: Bayes at Bletchley Park. Significance (Oxford, England), 7(2), 76–80. https://doi.org/10.1111/j.1740-9713.2010.00424.x

21. Thiese, Arnold, Z. C., Walker, S. D. (2015). The misuse and abuse of statistics in biomedical research. Biochemia Medica, 25(1), 5–11. https://doi.org/10.11613/BM.2015.001

22. Watson, M. D., Mesmer, B. L., Farrington, P. A. (2020). Engineering Statistics. In Engineering Elegant Systems: Theory of Systems Engineering (pp. 95-103). Huntsville, AL: NASA. doi:https://www.nasa.gov/sites/default/files/atoms/files /nasa_tp_20205003644_interactive2.pdf

23. Web of Science - Clarivate. (2023, March 01). Retrieved March 01, 2023, from https://www-webofscience-com.libproxy.uoregon.edu/wos/woscc/analyze-results/ afa19d59-de84-41d1-87c1-c67dfbe158a4-7658fe37