

Objections to Bayesian statistics

Andrew Gelman

Department of Statistics and Department of Political Science
Columbia University

Abstract. Bayesian inference is one of the more controversial approaches to statistics. The fundamental objections to Bayesian methods are twofold: on one hand, Bayesian methods are presented as an automatic inference engine, and this raises suspicion in anyone with applied experience. The second objection to Bayes comes from the opposite direction and addresses the subjective strand of Bayesian inference. This article presents a series of objections to Bayesian inference, written in the voice of a hypothetical anti-Bayesian statistician. The article is intended to elicit elaborations and extensions of these and other arguments from non-Bayesians and responses from Bayesians who might have different perspectives on these issues.

Keywords: Foundations, Comparisons to other methods

1 A Bayesian’s attempt to see the other side

Bayesian inference is one of the more controversial approaches to statistics, with both the promise and limitations of being a closed system of logic. There is an extensive literature, which sometimes seems to overwhelm that of Bayesian inference itself, on the advantages and disadvantages of Bayesian approaches. Bayesians’ contributions to this discussion have included defense (explaining how our methods reduce to classical methods as special cases, so that we can be as inoffensive as anybody if needed), affirmation (listing the problems that we can solve more effectively as Bayesians), and attack (pointing out gaps in classical methods).

The present article is unusual in representing a Bayesian’s presentation of what he views as the strongest non-Bayesian arguments. Although this originated as an April Fool’s blog entry (Gelman, 2008), I realized that these are strong arguments to be taken seriously—and ultimately accepted in some settings and refuted in others.

I welcome elaboration of these points from anti-Bayesians, as well as additional arguments not presented here. I have my own answers to some of these objections but do not present them here, in the interest of presenting an open forum for discussion.

Before getting to the objections, let me quickly define terms. “Bayesian inference” represents statistical estimation as the conditional distribution of parameters and unobserved data, given observed data. “Bayesian statisticians” are those who would apply Bayesian methods to *all* problems. (Everyone would apply Bayesian inference in situa-

tions where prior distributions have a physical basis or a plausible scientific model, as in genetics.) “Anti-Bayesians” are those who avoid Bayesian methods themselves and object to their use by others.

2 Overview of the objections

The fundamental objections to Bayesian methods are twofold: on one hand, Bayesian methods are presented as an automatic inference engine, and this raises suspicion in anyone with applied experience, who realizes that different methods work well in different settings (see, for example, Little, 2006). Bayesians promote the idea that a multiplicity of parameters can be handled via hierarchical, typically exchangeable, models, but it seems implausible that this could really work automatically. In contrast, much of the work in modern non-Bayesian statistics is focused on developing methods that give reasonable answers using minimal assumptions.

The second objection to Bayes comes from the opposite direction and addresses the subjective strand of Bayesian inference: the idea that prior and posterior distributions represent subjective states of knowledge. Here the concern from outsiders is, first, that as scientists we should be concerned with objective knowledge rather than subjective belief, and second, that it’s not clear how to assess subjective knowledge in any case.

Beyond these objections is a general impression of the shoddiness of some Bayesian analyses, combined with a feeling that Bayesian methods are being oversold as an all-purpose statistical solution to genuinely hard problems. Compared to classical inference, which focuses on how to extract the information available in data, Bayesian methods seem to quickly move to elaborate computation. It does not seem like a good thing for a generation of statistics to be ignorant of experimental design and analysis of variance, instead becoming experts on the convergence of the Gibbs sampler. In the short-term this represents a dead end, and in the long term it represents a withdrawal of statisticians from the deeper questions of inference and an invitation for econometricians, computer scientists, and others to move in and fill in the gap.

3 A torrent of objections

I find it clearest to present the objections to Bayesian statistics in the voice of a hypothetical anti-Bayesian statistician. I am imagining someone with experience in theoretical and applied statistics, who understands Bayes’ theorem but might not be aware of recent developments in the field. In presenting such a persona, I am not trying to mock or parody anyone but rather to present a strong firm statement of attitudes that deserve serious consideration.

Here follows the list of objections from a hypothetical or paradigmatic non-Bayesian:

Bayesian inference is a coherent mathematical theory but I don’t trust it in scientific applications. Subjective prior distributions don’t transfer well from person to person, and there’s no good objective principle for choosing a noninformative prior (even if that

concept were mathematically defined, which it's not). Where do prior distributions come from, anyway? I don't trust them and I see no reason to recommend that other people do, just so that I can have the warm feeling of philosophical coherence.

To put it another way, why should I believe your subjective prior? If I really believed it, then I could just feed you some data and ask you for your subjective posterior. That would save me a lot of effort!

As Brad Efron wrote in 1986, Bayesian theory requires a great deal of thought about the given situation to apply sensibly, and recommending that scientists use Bayes' theorem is like giving the neighborhood kids the key to your F-16. I'd rather start with tried and true methods, and then generalize using something I can trust, such as statistical theory and minimax principles, that don't depend your subjective beliefs. Especially when the priors I see in practice are typically just convenient conjugate forms. What a coincidence that, of all the infinite variety of priors that could be chosen, it always seems to be the normal, gamma, beta, etc., that turn out to be the right choices?

To restate these concerns mathematically: I like unbiased estimates and I like confidence intervals that really have their advertised confidence coverage. I know that these aren't always going to be possible, but I think the right way forward is to get as close to these goals as possible and to develop robust methods that work with minimal assumptions. The Bayesian approach—to give up even trying to approximate unbiasedness and to instead rely on stronger and stronger assumptions—that seems like the wrong way to go.

In the old days, Bayesian methods at least had the virtue of being mathematically clean. Nowadays, they all seem to be computed using Markov chain Monte Carlo, which means that, not only can you not realistically evaluate the statistical properties of the method, you can't even be sure it's converged, just adding one more item to the list of unverifiable (and unverified) assumptions. Computations for classical methods aren't easy—running from nested bootstraps at one extreme to asymptotic theory on the other—but there is a clear goal of designing procedures with proper coverage, in contrast to Bayesian simulation which seems stuck in an infinite regress of inferential uncertainty.

People tend to believe results that support their preconceptions and disbelieve results that surprise them. Bayesian methods encourage this undisciplined mode of thinking. I'm sure that many individual Bayesian statisticians are acting in good faith, but they're providing encouragement to sloppy and unethical scientists everywhere. And, probably worse, Bayesian techniques motivate even the best-intentioned researchers to get stuck in the rut of prior beliefs.

As the applied statistician Andrew Ehrenberg wrote in 1986, Bayesianism assumes: (a) Either a weak or uniform prior, in which case why bother?, (b) Or a strong prior, in which case why collect new data?, (c) Or more realistically, something in between, in which case Bayesianism always seems to duck the issue.

Nowadays people use a lot of empirical Bayes methods. I applaud the Bayesians'

newfound commitment to empiricism but am skeptical of this particular approach, which always seems to rely on an assumption of “exchangeability.” In political science, people are embracing Bayesian statistics as the latest methodological fad. Well, let me tell you something. The 50 states aren’t exchangeable. I’ve lived in a few of them and visited nearly all the others, and calling them exchangeable is just silly. Calling it a hierarchical or a multilevel model doesn’t change things—it’s an additional level of modeling that I’d rather not do. Call me old-fashioned, but I’d rather let the data speak without applying a probability distribution to something like the 50 states which are neither random nor a sample.

Also, don’t these empirical and hierarchical Bayes methods use the data twice? If you’re going to be Bayesian, then be Bayesian: it seems like a cop-out and contradictory to the Bayesian philosophy to estimate the prior from the data. If you want to do multilevel modeling, I prefer a method such as generalized estimating equations that makes minimal assumptions.

And don’t even get me started on what Bayesians say about data collection. The mathematics of Bayesian decision theory lead inexorably to the idea that random sampling and random treatment allocation are inefficient, that the best designs are deterministic. I have no quarrel with the mathematics here—the mistake lies deeper in the philosophical foundations, the idea that the goal of statistics is to make an optimal decision. A Bayes estimator is a statistical estimator that minimizes the average risk, but when we do statistics, we’re not trying to “minimize the average risk,” we’re trying to do estimation and hypothesis testing. If the Bayesian philosophy of axiomatic reasoning implies that we shouldn’t be doing random sampling, then that’s a strike against the theory right there. Bayesians also believe in the irrelevance of stopping times—that, if you stop an experiment based on the data, it doesn’t change your inference. Unfortunately for the Bayesian theory, the p-value *does* change when you alter the stopping rule, and no amount of philosophical reasoning will get you around that point.

I can’t keep track of what all those Bayesians are doing nowadays—unfortunately, all sorts of people are being seduced by the promises of automatic inference through the “magic of MCMC”—but I wish they would all just stop already and get back to doing statistics the way it should be done, back in the old days when a p-value stood for something, when a confidence interval meant what it said, and statistical bias was something to eliminate, not something to embrace.

Bibliographic note

I will not attempt to review here the literature on statistical objections to Bayesian inference. Some of the key issues were aired in the discussion of Lindley and Smith’s 1972 article on the hierarchical linear model. In the decades since this work and Box and Tiao’s and Berger’s definitive books on Bayesian inference and decision theory, the debates have shifted from theory toward practice. But many of the fundamental disputes remain and are worth airing on occasion, to see the extent to which modern developments in Bayesian and non-Bayesian methods alike can inform the discussion.

- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, second edition. New York: Springer-Verlag.
- Box, G. E. P., and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. New York: Wiley Classics.
- Efron, B. (1986). Why isn't everyone a Bayesian? *American Statistician* **40**, 1–5.
- Ehrenberg, A. S. C. (1986). Discussion of Racine, A., Grieve, A. P., Fluhler, H., and Smith, A. F. M., “Bayesian methods in practice: experiences in the pharmaceutical industry.” *Applied Statistics* **35**, 135–136.
- Gelman, A. (2008). Why I don't like Bayesian statistics. *Statistical Modeling, Causal Inference, and Social Science* blog, 1 April.
- Lindley, D. V., and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society B* **34**, 1–41.
- Little, R. J. (2006). Calibrated Bayes: a Bayes/frequentist roadmap. *American Statistician* **60**, 213–223.

Acknowledgments

The National Science Foundation, the National Institutes of Health, and the Columbia University Applied Statistics Center provided financial support for this work. The author also thanks Hal Stern for helpful comments and Brad Carlin for helpful comments and for setting up the discussion of this article.