Misconceptions about Bayesian modeling and analysis[1]

Andrew Gelman[2]

27 July 2010

Abstract. Bayesian inference is sometimes characterized as inappropriate for scientific research because of its subjectivity and dependence on models. We examine two versions of this argument, coming from the fields of economics and ecology, and discuss how these criticisms are aimed at a particular philosophy and set of methods that do not correspond to more modern methods of Bayesian data analysis. We hope to increase the awareness among the broader audience of statisticians and users of statistics of a view of Bayesian inference that is centered on models and model-checking rather than on subjective probability.

This is not another article expounding upon the superiority of the Bayesian approach to statistics. We take it as given that Bayesian methods are more useful in some settings than others, and in many situations it doesn't matter what you call your method as long as it works.[3] Our main concern is that people might be convinced by general anti-Bayesian arguments to avoid using statistical methods that would be effective for their purposes. The corresponding worry in the opposite direction is that pro-Bayesian advocacy will lead people to inappropriate methods.

We will discuss both these issues in the present article, which is an attempt by a Bayesian to communicate with anti-Bayesians. It is my impression that critics are often fighting an imaginary Bayes, or, at the very least, not the best and most up-to-date version of Bayesian inference. We applaud the denunciations of bad Bayesian ideas but we think the critics are misinformed in their attacks on Bayesian inference in general.

We discuss two sets of thoughtful criticisms, coming from economics (DiNardo, 2008) and ecology (Dennis, 2001, Lele and Dennis, 2009). If respected, empirically-minded researchers in these two quite different disciplines are unaware of modern Bayesian methods,[4] then it's a pretty

---

[1] We thank the Institute of Education Sciences, Department of Energy, National Science Foundation, and National Security Agency for partial support of this work.

[2] Department of Statistics and Department of Political Science, Columbia University, gelman@stat.columbia.edu

[3] For example: the posterior distributions involved in Bayesian hierarchical models can be considered as predictive distributions in a classical repeated-sampling framework; shrinkage of noisy estimates can be described as Bayesian or, more generally, as regularization; prior information is introduced into experimental design via power analysis; different sources of information can be combined using classical measurement-error models, and so forth.

[4] DiNardo does recognize the diversity of Bayesian approaches:

> In what follows, when I [DiNardo] describe something as "Bayesian" I do not mean to suggest any writer in particular holds all the views so attributed here. There is considerable heterogeneity: some view concepts like "the weight of evidence" as important, others do not. Some view expected utility as important, other do

good guess that confusion holds in many other quarters as well, so it seemed worth trying to spread the news of some important conceptual developments in Bayesian statistics that have taken place in the past two decades.

Brief summary of anti-Bayesian arguments

In our view, the key criticisms of Bayesian inference are:

1. Subjectivity
2. The prior distribution.

Regarding the first point, it seems like a step backwards to introduce subjectivity into scientific research, which is among the most objective of human endeavors. The second point is related but slightly different: Even if the prior distribution could be assigned on an objective basis, anti-Bayesians ask why one would want to add an additional assumption to a statistical model if it is not needed.

Along with these comes a more sociological criticism:

3. Arrogance.

Bayesians often seem to present themselves as having all the answers, of working in a perfectly logical framework. If you do not already use Bayesian methods, it can be annoying to hear people say that your approach is logically inferior. See Gelman (2008) for a recent discussion of criticisms of Bayesian inference.

The standard Bayesian responses to the above critiques are: (1) all scientific inference is inherently subjective (for example, in the choice of variables to include in a model); (2a) the prior distribution can add useful information; (2b) in any case, inference typically depends much more on the likelihood function than on the prior distribution, and so if you want to avoid models, you should start by avoiding likelihoods.[5]

We do not claim that the above paragraph is in any sense a coherent defense of Bayesian methods; it merely sketches an outline of a defense that we feel is best done not with axiomatics but through examples (see Gelman et al., 2003). Our real point here is that standard anti-Bayesian attitudes are centered on a version of Bayesian inference that is far different from what we do in our statistical analyses.

> not. This is not intended to be a "primer" on Bayesian statistics. Neither is it intended to be a "critique" of Bayesian views. . . . My purpose is not to do Bayesian ideas justice (or injustice!) but rather, to try selectively choose some implications of various strands of Bayesianism and non-Bayesianism for actual statistical practice that highlight their differences so as to be clear to a non-Bayesian perspective.

His reference to "actual statistical practice," however, suggests that he is trying to discuss the Bayesian methods that are used in applications. What I suspect is that DiNardo is simply *unaware* of the modern approach to Bayesian data analysis which is based on modeling and active model checking ("severe testing," to use the phrase of Mayo, 1996).

[5] There are some statistical methods that are (or attempt to be) model-free. The extent to which this is possible has been debated (George Box, for example, has argued that many so-called nonparametric models end up relying heavily on assumptions of independence, additivity, and distributional invariance; see DeGroot, 1987, p.248), but for our purposes here we are focusing on criticisms of Bayesian methods rather than on criticisms of statistical models in general.

Changing times

During the past few decades, the context of these criticisms, and the nature of anti-Bayesianism, has changed. Bayesian statisticians were formerly considered to be something of a nuisance, a buzzing colony of theoreticians who occasionally offered a superficially attractive optimality argument but could be safely ignored when it came to statistical practice. More recently, however, we have started to hear the opposite argument, that Bayes is all too easy, with button-pushing replacing thought.

For example, as late as 2001, ecologist Brian Dennis described the Bayesian statistical literature as "highly mathematical, and a scientist is right to question whether the attraction is mathematical instead of scientific. . . . the numerical methods [for Bayesian inference] at present involve heavy computer programming efforts, post-calculus statistics knowledge, and sometimes days of computer time; the methods are not ready yet for routine use by busy laboratory or field scientists."

Only eight years later, this same author (in collaboration with statistician Subhash Lele) wrote, "Unfortunately expositions of Bayesian methods for hierarchical models have tended to emphasize practice while deemphasizing the inferential principles involved."

As applied Bayesians, we welcome this transition but it does not dislodge the more fundamental criticisms regarding subjectivity and prior distributions.

Bayesian statistics as it is seen from the outside vs. Bayesian data analysis as it is seen from the inside

We think of Bayesian inference as a generalization of least squares and maximum likelihood, with prior distributions and multilevel models (which can also be viewed as simultaneous equations or measurement-error models) as a way of regularizing or obtaining more stable estimates. Public statistics performed this way can be more effective than classical estimates that commonly need to be interpreted with one eye on the sample size.

We consider several areas in which we feel that common views about Bayesian statistics have not caught up with best practices.

Subjective prior distributions. We agree with Bernardo (2008) that Bayesian priors are objective in the same sense that classical methods are, "in that the final result only depends on the model assumed and the data obtained." Whether our methods are classical or Bayesian, it is a good idea to use models that make sense, to consider reasonable alternatives, and to check our assumptions where possible. Like all science, statistics is made of subjective procedures that yield objectively testable results.

More specifically, we treat any numbers used in prior distributions as data, subject to the same rules of evidence that apply to data used in the likelihood. Dennis (2001) discusses the malign impact of an ill-informed prior distribution on a hypothetical ecological analysis. But this sort of problem can occur in a classical analysis too—once you allow the analyst to include unverified data of questionable provenance. In a traditionally fully-subjective Bayesian world, sure, this is possible. But modern Bayesian analysis has no place for arbitrary numbers that come with no scientific justification.

Expert opinions. Lele and Dennis (2009) write:

> We do believe that expert opinion can and should be brought into ecological analyses . . . although not by Bayesian methods. Recently, Lele and Allen (2006) showed how expert opinion can be incorporated using a frequentist framework by eliciting data instead of a prior. The methodology . . . automatically weighs expert opinion and hard data according to their Fisher information about the parameters of the process.

This is fine—but we would consider this to be a perfectly reasonable Bayesian measurement-error model. To a modern Bayesian, there is no sharp distinction between “likelihood” and “prior distribution,” and it makes perfect sense to treat expert opinions as data and model them as such.

Experimental design. It is commonly stated that randomization is not allowed under Bayesian principles. For example, DiNardo writes that, according to Bayesians,

> The absence or presence of data mining strategies, specification mining, non-random sampling, non-random assignment are (should be) irrelevant to the inference of a set of data. Put differently, what could have happened but didn’t in an experiment should make no difference to the evidential import of the experiment.

He supports this claim with quotations from 1971 and 1976. But Bayesian inference has advanced a lot in the past thirty-five years. In particular, it is well-understood that a correct Bayesian analysis must condition on all the information used in the selection of data. (See chapter 7 of Gelman et al., 2003,[6] for a discussion of this principle with several examples, including censored and truncated data and violation of simple random sampling and random assignment in surveys and experiments.)

Likelihood principle. In a Bayesian analysis, the data enter the posterior distribution only through the likelihood; thus issues such as sequential data collection seem not to come into play. Actually, though, data collection *is* relevant to Bayesian data analysis: it enters in the model-checking stage, when observed data are compared to replications from the posterior predictive distribution. Performing these replications requires simulation of the data-collection process, and model checks under sequentially-collected data, for example, will look different from model checks for data collected using a fixed sample size. (For a simple example, see exercise 6.6 of Gelman et al., 2003.)

Model checking. If a Bayesian model truly represents a subjective belief, than perhaps it is indeed untestable, as is sometimes alleged. In the modern view of Bayesian data analysis,

[6] We cite our own writings because that's what we are most familiar with. It should be easy enough to form a similar reply using the words of others.

however, a model is not a belief but is rather a set of assumptions, and these assumptions can indeed be checked by comparing to data (see Gelman et al., 2003, chapter 6).

Arrogance. DiNardo writes that Bayesians believe that "The problem of 'how to reason' has been solved." No. A Bayesian inference is only as good as its models. Bayesians are busy developing, applying, checking, and extending new classes of models. Consider, for example, the explosion of research in nonparametric Bayes in recent years. Given that "how to reason Bayesianly" includes model specification, and given that model specification is wide open, it would be foolish to claim that the problem of reasoning has been solved, and applied Bayesians do not make that claim.

The arrogance of many Bayesian statisticians has been well documented over the years. But, more than anything else, this may be a product of scientists treating statisticians as authority figures. Plenty of non-Bayesian statisticians are arrogant too.

Philosophical differences. We view the Bayesian approach is a useful way to come up with an estimator in complicated problems with structured data. It is perfectly fine (and, in fact, often recommended by Bayesians; see, for example, Box, 1980, and Rubin, 1984) to study the statistical properties of such estimators.

A way forward

DiNardo, Dennis, and others are hardly to blame for their negative impressions of Bayesian statistics. In fact, one could plausibly argue that they are correct and that ours is the minority view. For example, the Wikipedia article on Bayesian inference currently begins:

> Bayesian inference is statistical inference in which evidence or observations are used to update or to newly infer the probability that a hypothesis may be true. Bayesian inference uses aspects of the scientific method, which involves collecting evidence that is meant to be consistent or inconsistent with a given hypothesis. As evidence accumulates, the degree of belief in a hypothesis ought to change. With enough evidence, it should become very high or very low. . . . Bayesian inference uses a numerical estimate of the degree of belief in a hypothesis before evidence has been observed and calculates a numerical estimate of the degree of belief in the hypothesis after evidence has been observed. . . . Bayesian inference usually relies on degrees of belief, or subjective probabilities, in the induction process and does not necessarily claim to provide an objective method of induction.

Whether or not Wikipedia is an authority, it is often taken as such, and whether or not it represents a majority opinion (to the extent possible given the potential diversity of Bayesian approaches; see Good, 1971), it cannot be ignored.

In his extensive discussion of the philosophy of inference, DiNardo remarks, perhaps accurately, that the literature on the philosophy of statistics is dominated by Bayesians with extreme and often nutty views. And this frustrates him. But we suspect he is not looking in the right place.

When we wrote our book on Bayesian data analysis, we were careful *not* to include the usual philosophical arguments that were at that time considered standard in any Bayesian presentation. We decided to skip the defensiveness and just jump straight to the models and the applications. This worked well, I think, but it has led people without direct experience in applied Bayesian statistics to not notice the implicit philosophical content that is contained implicitly in the methods.[7]

And this is the paradox of philosophizing. If we *had* put 50 or 100 pages of philosophy into our book (rather than discussing model checking, randomization, the limited range of applicability of the likelihood principle, etc., in separate places in the book), that would've been fine, but then we would've diluted our message that Bayesian data analysis is about what works, not about what is theoretically coherent. Many people find philosophical arguments to be irritating and irrelevant to practice. Thus, to get to the point, it can be a good idea to avoid the philosophical discussions. But, as the saying goes, if philosophy is outlawed, only outlaws will do philosophy**.** DiNardo is responding to the outlaws. I hope this article will alert him and others to the philosophy that is all around him every day.

We agree with the critics that Bayesians are doing themselves no favors by describing their inferences as subjective and personalistic, and we agree that there is no meaningful way to construct noninformative priors in general, any more than there is a way to construct noninformative, objectively-defined likelihoods or functional relationships in most settings.

However, it turns out that the mathematics of probability works well in combining information from multiple sources, hence the move from “subjective Bayes” to “hierarchical Bayes,” and the shift from the idea of Bayes as merging subjective opinions with data to the idea of Bayes as an approach for combining data from multiple sources—a task that is important in economics, ecology, and many other sciences.

We agree with DiNardo that “what type of questions one is interested asking often suggests what type of statistics one finds useful,” and we have not sought here to convert anyone to Bayeisanism. But if we have persuaded some critics to update their views on what Bayesian data analysis is and could be,[8] we would consider this article to be a success.

References

Bernardo, J. (2008). Discussion of “Objections to Bayesian statistics.” *Bayesian Analysis* 3.

Box, G. E. P. (1980). Sampling and Bayes inference in scientific modelling and robustness. *Journal of the Royal Statistical Society* A 143, 383-430.

---

[7] Here are two examples of implicit philosophy: (1) the routine use of hierarchical models shifts much of the burden of specifying prior distributions away from elicitation of experts and toward data modeling; (2) model checking based on simulation of replicated data implies the data-collection process is indeed relevant to Bayesian data analysis.

[8] In particular, we point interested non-Bayesians to the book of Carlin and Louis (1996), who discuss foundational issues in an applied Bayesian context.

Carlin, B. P., and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*, second edition. London: Chapman and Hall.

DeGroot, M. (1987). A conversation with George Box. *Statistical Science* 2, 239-258.

Dennis, B. (2001). Statistics and the scientific method in ecology. In *The Nature of Scientific Evidence*, ed. M. L. Taper and S. R. Lele. University of Chicago Press.

DiNardo, J. (2008). Introductory remarks on metastatistics for the practically minded non-Bayesian regression runner. In *Palgrave Handbook of Econometrics: Vol. 2. Applied Econometrics*, ed. T. C. Mills and K. Patterson. Basingstoke: Macmillan. http://www-personal.umich.edu/ jdinardo/draftfinal4.pdf

Gelman, A. (2008). Response to discussions of "Objections to Bayesian statistics." *Bayesian Analysis* 3, 468-478.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). Bayesian Data Analysis, second edition. London: CRC Press.

Gelman, A., and Shalizi, C. (2010). Philosophy and the practice of Bayesian statistics. Technical report, Department of Statistics, Columbia University.

Good, I. J. (1971). 46656 varieties of Bayesians. *American Statistician* 25, 62-63.

Lele, S. R., and K. L. Allen. 2006. On using expert opinion in ecological analyses: A frequentist approach. *Environmetrics* 17, 683-704.

Lele, S. R., and Dennis, B. (2009). Bayesian methods for hierarchical models: Are ecologists making a Faustian bargain? *Ecological Applications* 19, 581-584.

Ponciano, J. M., Taper, M. L., Dennis, B., and Lele, S. R. (2009). Hierarchical models in ecology: Confidence intervals, hypothesis testing, and model selection using data cloning. *Ecology* 90, 356-362.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* 12, 1151-1172.