The humble Bayesian: model checking from a fully Bayesian perspective

Richard D. Morey and Jan-Willem Romeijn
University of Groningen

Jeffrey N. Rouder
University of Missouri

Corresponding Author:
Richard D. Morey
DMPG
Grote Kruisstraat 2/1
9712TS Groningen, The Netherlands
`r.d.morey@rug.nl`

Abstract

Gelman:Shalizi:inpress criticize what they call the "usual story" in Bayesian statistics: that the distribution over hypotheses or models is the sole means of statistical inference, thus excluding model checking and revision, and that inference is inductivist rather than deductivist. They present an alternative hypothetico-deductive approach to remedy both shortcomings. We agree with ?'s criticism of the usual story, but disagree on whether Bayesian confirmation theory should be abandoned. We advocate a humble Bayesian approach, in which Bayesian confirmation theory is the central inferential method. A humble Bayesian checks her models and critically assesses whether the Bayesian statistical inferences can reasonably be called upon to support real-world inferences.

Address correspondence to: r.d.morey@rug.nl.

Modern statistics is a diverse field with disagreements even about the most basic and foundational issues. Savage:1972 noted 60 years ago that there were scarcely any accepted facts about the foundations of statistics, and ?'s ((?)) article (hereafter G&S) is proof that disagreements exist even today. But G&S also reveal that in spite of these disagreements, or perhaps rather because of them, statisticians continue to develop useful ways of learning from data.

Comparison with G&S

We agree with G&S's critique of what they call the "usual story" in Bayesian statistics, and also acknowledge the usefulness of the procedures they advocate. But rather than abandon the traditional Bayesian framework, we promote a perspective on Bayesian statistics that is strengthened through the use of model checking procedures.

Overconfidence is wrong, but Bayes is right

G&S introduce the usual story of Bayesian data analysis: that all information necessary for inference is contained in Bayesian quantities such as posterior distributions or model posteriors. In this story, model checking is not performed at all; posterior quantities tell us the relative plausibilities of parameter values or models. A Bayesian of this stripe is what we refer to as an "overconfident" Bayesian. To our mind, the overconfident Bayesian is an extreme point in the spectrum of Bayesians. We ourselves routinely perform model checks in our own work Morey:etal:2009,Morey:etal:2008a,Rouder:etal:2008d and we believe that most practicing Bayesian statisticians worry about the appropriateness of their models and hence engage in model checking.

One reason for the impression that overconfident Bayes features prominently in the philosophy of statistics may be that, in philosophy, Bayesian inference is often considered as part of a logic DeFinetti:1995,Howson:2001,Romeijn:2011. Philosophers of statistics focus on the correctness of the inferential step rather than on the truth or falsity of the premises. In other words, the focus lies on the Bayesian data analysis and not on the appropriateness of the model. However, as indicated by the parallel interest in model selection among philosophers of statistics Kieseppa:2001,Forster:Sober:1994,Romeijn:vandeSchoot:2008,Romeijn:etal:2012, the focus on correctness should not be taken to indicate that, according to philosophers of statistics, valid inference is all there is to good statistical practice.

In contrast to overconfident Bayesianism, the scheme that G&S propose for model checking is not Bayesian. Their view is what they call hypothetico-deductive or, in other places, falsificationist: models are judged by how well they accommodate the data and then retained or discarded. The core of the view seems to be that model checking is not regulated by an inductive, but rather by a deductive mode of inference. Statistical models entail probabilistic empirical consequences, and they do so deductively, as a matter of mathematical fact. These probabilistic consequences can then be compared to data to arrive at a judgment on the model.

We accept that model checking is an integral part of good statistical practice. The overconfident Bayesian is wrong. But we believe that if model checking is to become

a primary method of statistical inference, more detail is needed on how it is supposed to be done. In other words, we require a theory of inference using model checking. Although G&S offer a number of tools and procedures and a general philosophy, they do not offer a theory of inference. But a suitable theory of inference already exists: the Bayesian confirmatory framework. A reasonable Bayesian can use model checking alongside traditional Bayesian analyses, casting the model checking itself in a Bayesian light.

Conceptual issues for G&S

G&S briefly discuss arguments for the Bayesian consistency and rationality but they do not seem pursuaded. To our mind, it is a major advantage of a Bayesian approach to model checking that it inherits the conceptual clarity and coherence of Bayesian theory generally. We provide some detail on Bayesian model checking below. Here we note two conceptual issues for G&S.

As G&S indicate, model checking typically proceeds by finding out that the model under scrutiny is false, as its empirical consequences do not match the data. Strictly speaking statistical models can of course not be falsified, since probabilistic consequences cannot be contradicted by data. Much like Mayo:1996,Mayo:Spanos:2011, it seems that G&S speak of falsificationism and deductivism by proxy: highly improbable data are somehow considered close enough to impossible data to effect a form of falsification. For philosophers and statisticians who champion the validity of the inferences this attitude is somewhat puzzling, especially since a valid inferential framework is already available in Bayesian theory. Now it may be that G&S are simply not bothered by these concerns, but we think they should be, and that the broad strokes in which G&S's deductivism is painted need to be revisited with a finer brush.

Furthermore, G&S's abandonment of the Bayesian framework has consequences for their proposed method of model checking. G&S imply that they have abandoned the Bayesian framework even to the extent of rejecting a probabilistic interpretation of the Bayesian prior, which to them is "more like a regularization device, akin to the penalization terms added to the sum of squared errors when doing ridge regression and the lasso or spline smoothing." This rejection, however, has consequences. The probabilistic interpretation of the posterior arises from the probabilistic interpretation of the prior. Abandoning the probabilistic interpretation of the prior threatens the interpretation of the corresponding posterior, and thus the interpretation of the posterior predictive p values[1]. Since posterior predictive p values are one of the primary methods G&S have advocated for model checking, it is important that these p values be interpretable.

Summing up, we largely agree with G&S in their dislike of overconfident Bayes and on the importance of model checking. But we feel that G&S need to provide a

---

[1]One could argue that the use of improper priors also threatens the interpretation of Bayesian quantities as well, since they don't have a ready probability interpretation. However, many improper priors are limits of proper priors. The interpretation of the prior in this case is quite different from a nonprobabilistic "regularization device". The debate over the use of improper priors is ongoing and interesting Berger:2006,Goldstein:2006, but we do not wish to engage it here.

theory. G&S's approach compares unfavourably to the coherence and conceptual clarity of Bayesianism.

Departing from G&S

Apart from being Bayesian, our perspective differs from G&S in two ways: one pragmatic, and the other philosophical. Pragmatically, we believe that although G&S's approach is likely to be successful for the types of problems they encounter, it is not ideal for questions we commonly encounter. Philosophically, we disagree with G&S that all models are wrong.

The importance of invariances

Statistics is such a diverse field in part because of the wide variety of questions that statistics is required to address. Differences in goals and applications lead to immediate differences in statistical philosophy. G&S state, "The statistician begins with a model that stochastically generates all the data y, whose joint distribution is specified as a function of a vector of parameters θ from a space Θ (which may, in the case of some so-called non-parametric models, be infinite dimensional)." In contrast, we start from a theoretical question about a scientific phenomena of interest. We almost always interested in assessing invariances: those elements of structure or constancy in a complex relationship among variables. A classic example of invariances are Kepler's laws of planetary motion. Although the trajectories of the planets seem complicated when viewed from Earth, Kepler was able to deduce a set of three simple constraints that governed the relationship among the observables. The search for simplifying structure is ubiquitous in the the natural sciences, and in many experimental social sciences Morey:Rouder:2011,Rouder:Morey:2011a,Rouder:etal:2008a. Theoretical differences over models amount to different constraints on data.

For problems like the ones G&S handle, which are often involve linear models (or generalized linear models) of varying complexity, the model checking approach is feasible. However, for many of the questions we encounter, it is difficult to imagine how model checking could serve as the primary mode of inference. If two theoretical positions are represented by radically different stochastic models (as will often be the case in psychology), both models will likely misfit in different ways, and it may not be obvious how to compare the two. We seek methods for either moving toward less-complex models embedding more invariances, or across classes of models embedding different invariances. The traditional Bayesian framework offers a natural way of answering the questions we face, with the benefit that it comes with a formal inferential framework.

Models are neither true nor false

G&S focus on the issues of "false" models in statistical inference, but we believe the idea of "false" models to be unhelpful. As representations, scientific models, including statistical models, are neither true nor false Bailer-Jones:2003,Hughes:1997,Hutten:1954, unlike the propositions about the world that they represent. We believe that ?'s ((?))

Figure 1. ?'s ((?)) DDI model of scientific representation. "World" squares represent phenomena (dark) or propositions about phenomena (light); "Representation" squares represent models (dark) or inferences about models (light).

famous dictum, "all models are false but some are useful," could be shortened to "some models are useful" without any loss. The main question to us is to what extent the inferences made using representations can be applied to corresponding inferences about the world. Useful statistical models and procedures will provide inferences that can be "interpreted" in such a way as to be useful for inference in the real world.

Hughes:1997 describes a framework which can be used to understand the process of how scientific models are used, which he called the denotation, demonstration, interpretation (DDI) framework (Figure **??**). To help answer the researcher's question, the statistician will develop a statistical model. This move from the "real world" into the model world Hughes calls denotation. The statistical model, by necessity, is an idealized representation. The researcher's hypotheses about the real world are not answered directly; instead, questions about parameters of the statistical model are answered. Inference about the mean value of a population, for instance, is replaced by inference about the normal distribution, which is a representation of the population of interest. Hughes called this process of acting on representations demonstration. Finally, inferences with respect to the representation must be translated back into the world through interpretation of the statistical inference. Hughes conception of the role of models is central how we view Bayesian analysis.

The humble Bayesian

With these differences in perspective in place, we now spell out how the practice of model checking can be aligned with Bayesian inference, as long as we are suitably humble in applying our inferences. We call our view "humble Bayes[2]," but we make no claims as to its novelty. We note that G&S's list of those who advocate some form of model checking is a veritable who's-who of 20th century Bayesian statisticians, and we suspect that most Bayesians adhere to a similar philosophy, without giving it a name.

Open-minded inference

In the foregoing we noted that G&S's falsificationism is not easily incorporated in a coherent theory of model checking. But for our Bayesian theory, we borrow from falsificationism what we take to be its greatest virtue: its open-mindedness. To its credit, there is no suggestion in the approach of G&S that the models presently under consideration are in some sense true, and new models can enter the arena at any stage of investigation. By contrast, a Bayesian who is pondering over a fixed set of models seems ultimately closed-minded. She has a prior probability over the set which expresses her belief in each

---

[2]Readers who have interacted with Bayesians may find the term "humble Bayesian" oxymoronic.

of the options available, and these priors add up to one, meaning that the disjunction of the models is believed with absolute certainty [cf.][]Dawid:1982.

If, on the other hand, we decide to employ odds as expressions of relative belief[3], then it is left open whether or not the probabilities of the models under consideration sum to unity. A Bayesian who employs odds is silent on whether or not she is in possession of the true model, and, in fact, needn't acknowledge the existence of a true model at all. But such a Bayesian is nevertheless able to incorporate prior beliefs into the inference. With minor interpretative adjustments, it is possible to incorporate openmindedness in the Bayesian inferential framework.

The primary inferential machinery in humble Bayesianism is thus traditionally Bayesian, using posterior distributions, model odds, and Bayes factors, the choice of which is largely driven by the research question. These Bayesian quantities are used to perform inferences within the statistical models at hand, but also to evaluate models and compare them to one another. This is unlike the overconfident Bayesian, simply because the models are questioned. Model checking serves two roles, which we can spell out in terms of the perspective on models given above: determination of the extent to which inferences can be carried from the statistical representation into the real world, and support of the generation of new models for comparison.

Model checking assures applicability of Bayesian inferences

In scientific settings, the quantities of interest are not quantities in any statistical model; rather, a researcher has questions about a particular population or process. These questions, if they are well-formed, can be reframed in terms of propositions about the world that are either true or false. There are uncountably many statistical models that could be used to help answer the researcher's questions, but we emphasize that the original question is not itself a question about a statistical parameter or model.

With this understanding, Bayesian confirmation theory still provides meaningful inferences. The goal of humble Bayesian confirmation theory is not to confirm a "true" model. Because the model itself is not true (nor is it false), neither confirming it nor falsifying it can be our goal. However, we can take a confirmation as indicating something important about the world. G&S mention that Bayes factors and posterior probabilities can be useful as long as they "are not taken too seriously." Our reason for not taking them too seriously is not that the underlying models are false; rather, it is that they are not the ultimate target for inference. The Bayes factor or posterior probability must be interpreted. If models are useful, statements about statistical parameters will correspond to statements about the world, but this correspondence will not be exact.

Overconfident Bayes is problematic because it lacks the necessary humility that accompanies understanding that inferences are based on representations. We agree that there is a certain silliness in computing a posterior odds between model A and model B, seeing that it is in favor of model A by one-million to one, and then declaring that model A has a 99.9999% probability of being true. But this silliness arises not from model A being

---

[3]"Belief" in this sense may be part of the statistical representation, and may or may not reflect the analysts belief in the corresponding proposition in the real world.

false. It arises from the fact that the representation of possibilities is quite likely impoverished because there are only two models. This impoverished representation makes translating the representational statistical inferences into inferences pertaining to the real world difficult or impossible[4]. For this reason, we prefer to speak of "model comparison" rather than "model selection": Models needn't ever be selected as true, but they can be compared in meaningful and informative ways.

The key, then, is to ensure that our statistical inferences can be interpreted in a useful way into real-world inferences. What must we do to ensure that demonstrations in the representation realm can be interpreted in such a way that they correspond in a useful way to statements about the world? This is a difficult question to answer, but at minimum, we believe it requires that the ancillary assumptions used to generate models are not unreasonable. Even in a fully Bayesian framework, model checks are necessary. Model checks help to assure ourselves that interpretation of the results is possible in a way that is useful for real-world inferences.

Model checking helps generate new models

In addition to helping assure ourselves that our Bayesian quantities are useful, model checks also spur the creation of new models, which can then be tested within the standard Bayesian model testing framework. G&S describe model testing as being outside the scope of Bayesian confirmation theory, and we agree. This should not be seen as a failure of Bayesian confirmation theory, but rather an admission that Bayesian confirmation theory cannot describe all aspects of the data analysis cycle. It would be widely agreed that the initial generation of models is outside of Bayesian confirmation theory; it should then be no surprise that subsequent generation of models is also outside its scope.

Generating multiple models in Hughes' denotation phase allows for a richer representation of the world. Because the quality of our inferences are related to the richness of our representation of the world (or possible worlds), generation of new models is essential to ensuring that our Bayesian inferences are applicable to real-world scenarios. Statistical inferences, including Bayesian ones, are only as useful as the underlying representation admits.

We therefore believe that model checking complements the Bayesian confirmatory approach to statistical inference. To a humble Bayesian, models are not true or false, but are representations. The humility in the humble Bayesian approach comes from understanding that these models are not the ultimate target of inference, and that model checking helps to ensure that we can bridge the gap between the representational world and the real world.

Conclusion

The humble Bayesian approach we have sketched out here has the advantage that it retains the core of the Bayesian confirmatory method with its formal inferential theory,

---

[4]On the other hand, in some research scenarios inference from impoverished representations may be possible. The usefulness of a two-model comparison is highly dependent on the phenomenon and research question.

something that G&S's approach lacks. It avoids many of the criticisms of G&S by keeping an open mind through model checking, and through humility, by understanding that Bayesian quantities must be interpreted. The applicability of Bayesian quantities will be determined by the quality of the statistical representation, which can be checked using the methods G&S advocate.