Bayesian Analysis (2006) **1**, Number 3, pp. 465–472

Subjectivity and objectivity in Bayesian statistics: rejoinder to the discussion

Michael Goldstein*

It has been very interesting to engage in this discussion on subjective and objective issues in Bayesian statistics. I hope that we have generated more light than heat. Thanks to all the discussants and *Bayesian Analysis* for providing the forum. Here are some reactions to the discussion.

1 Comments on Jim Berger's paper

Jim has provided an excellent overview of the considerations that lead to the "objective" Bayes position, emphasising the pragmatic advantages in simplifying the formulation and resulting analysis for the Bayesian statistician, who, after all, has a hard enough job just carrying out any version of a Bayesian analysis for substantial problems. I am happy to accept such simplifications in many situations, as I discuss in my consideration of pragmatic subjectivism. So, why does this issue generate so much controversy?

When discussing subjectivity and objectivity in Bayesian statistics, there are various issues that can easily become confused. There are deep philosophical questions as to what we mean by these terms and common sense considerations as to how these terms are generally used and understood in practice. These distinctions are mirrored by the corresponding methodological considerations as to what constitutes a good analysis in principle, and what is achievable in practice.

Let's consider the meaning of these terms. Here is how the The Internet Encyclopedia of Philosophy explains the distinction:

"Objective judgment or belief" refers to a judgment or belief based on objectively strong supporting evidence, the sort of evidence that would be compelling for any rational being. A subjective judgment would then seem to be a judgment or belief supported by evidence that is compelling for some rational beings (subjects) but not compelling for others.

[Objectivity, D.H. Mulder, The Internet Encyclopedia of Philosophy, http://www.iep.utm.edu/]

This seems reasonable to me. I think that it corresponds more or less to how the term is commonly used and understood and explains why it is valued as a gold standard in science and elsewhere. This is the view that I expressed in my article. A common aim for a scientific Bayesian analysis is to ascertain whether the data is sufficiently convincing that any reasonable assignment of prior judgments would lead to roughly the same conclusions. If we called such an outcome an "objective Bayesian analysis"

*Department of Mathematical Sciences, University of Durham, UK, http://fourier.dur.ac.uk:8000/stats/people/mg/mg.html

 ba0003

then I don't think that this would be grossly misleading. This is because the term "objective" attaches to the judgements that we reach and thus to the claims that we may make for the results of the analysis.

This is my fundamental objection to all of those approaches which advocate the use of the term "objective Bayes" to refer simply to the style of the analysis — i.e. one which lacks subjective elements. Is this what users want or think that they are getting? I very much doubt it. Users are happy with objective Bayes methods because they think that this approach provides objective judgements as commonly understood. How could they think otherwise? They are not experts in statistical reasoning! We, on the other hand, are meant to be expert, and we should be careful, straightforward and honest about this. If we want to run our data through a "neutral" prior specification for pragmatic reasons — and Jim gives many good reasons why we may want to do this — then there are many attributions which we may use which are not so dangerously misleading. I prefer the strictly neutral term "formal Bayes," simply because this clarifies that the prior is chosen to meet some formal property rather than a judgement of knowledge for the area. If the main appeal of the approach in a particular case is that it combines good frequency properties with the formal Bayes structure, then "frequentist Bayes" would serve.

Of course, in Jim's phrase about alternative choices of name, ".. none carries the simplicity or will carry the same weight outside of statistics as 'objective'." However this argument cuts both ways. If we claimed to offer a money back guarantee if our statistical analyses were in error, then this would be simple and carry a lot of weight in the user community. However, we can easily imagine the reaction from users who tried to claim the money back guarantee, only to be told that there was no actual guarantee of a refund, but that the Bayesian community had simply judged that to say that there would be a refund would carry a lot of weight outside of statistics. Sometimes, we can carry the "weight" of an objective analysis, and otherwise not and we should be clear about this.

Is the teaching of elementary statistics greatly simplified through the device of objective Bayes? This might be true if we take the view that our students will not understand what we tell them anyway, so we might as well tell them the simplest story that we can. I am a bit less pessimistic however. In my view, a first statistics course should explain that there are two types of properties that are valued by different groups of statisticians, namely long run frequency properties and statements of belief about individual cases. Having explained that these are in principle different, we may now quote (in more or less detail depending on the background of the students) those properties that ensure in many problems, for large samples, that the approaches are essentially in agreement. This explains the consensus obtained for large samples and offers up a wide range of problems that students can tackle without great inferential sophistication, while placing a clear marker that small sample inferences are more subtle and complex and require more thoughtful analysis.

2 Response to J. Andres Christen

Should we drop the qualifier "subjective" from the description of a Bayesian analysis? We certainly do need to have a better categorisation of the variety of possible Bayesian analyses that we might carry out; for example, a Bayesian analysis might describe the range of allowable judgements within a community of experts, or it might reflect the judgements of a single expert. If the latter, then the expert judgements might simply be stated as assumptions for the analysis or a careful attempt might be made to source the judgements and to convince the reader of their plausibility, and so forth. I tend to view the analysis of the expert community as a scientific Bayesian analysis, and the analysis of the individual as a personal Bayesian analysis, although if the prior assessments of the individual are carefully and convincingly justified then a personal Bayesian analysis can have great scientific value. Given such multiplicity, maybe the subjective qualifier is unhelpful; there are more important distinctions to be made.

3 Response to David Draper

These comments on the subjective aspects of calibration are very interesting. In the example, the effect follows from good subjective assessments over the relation between past and future trials. I view this as part of the general issue that injecting relevant external knowledge into the analysis is likely to improve overall long term performance. It is hard to formalise this notion, as it requires us to formalise the notion that what we consider to be relevant information actually is relevant and that we haven't made a mistake in choosing to include it. Subjectively, I am happy to believe that the things that I believe are more likely to be helpful than not, while keeping a watchful diagnostic eye to trap gross misjudgements.

As to the specific question about the equation $\mathrm{P}_t(A) = \mathrm{P}(A|\mathcal{B}) + R$, yes, there are constraints on R which derive from the constraints on each other quantity in the equation. I tend not to emphasise these merely because the basic relation is developed in unconstrained form relating conditional and posterior expectation, and the residual simply inherits whichever constraints follow in any particular application.

4 Response to Rob Kass

Good questions! Let's play. In my account, I have argued that we cannot really understand these issues without distinguishing two basically different types of Bayesian analysis.

Let us call a *fully Bayesian analysis* one which is concerned with a problem which is sufficiently important that it merits as careful an analysis as we can give, and for which we make every effort to do absolutely the best job that we can. Judgements underlying the prior and the model are carefully made and fully and explicitly justified. Where relevant, potential differences in the expert community are explicitly represented and reflected in the resulting sensitivity analysis. The overall specification has survived a

stringent diagnostic assessment. The subjectivist gold standard.

Similarly, call a *limited Bayesian analysis* one for which either the problem is not sufficiently important to carry out a full treatment, or we feel insufficiently skilled or resourced to follow such an approach, but where we still want to provide some helpful analysis for the problem.

(Of course, there is a continuum of intermediate positions. For example, Bayes linear analysis seeks to claim the most important features of the gold standard, while enjoying many of the simplifications of the pragmatic approach.)

So, to Rob's questions.

[1] For a fully Bayesian analysis, frequentist operating characteristics are irrelevant. We want to make the best assessment for the case at hand. Any conflict between good frequency rules and best subjective judgements should be resolved in favour of subjective assessments. Even in the long run, we will expect to do better by doing our best for each case. Besides, it would be absurd to decide that our most carefully assessed judgement is of one form but that we will make a quite different judgement simply to average out with a host of unrelated analyses.

For a limited Bayesian analysis, the situation is less clear. The subjectivist reason why we might be interested in frequentist operating characteristics is that they are preposterior valid for everybody. For example, if you tell me that you have calculated a 95% confidence interval for some quantity, without telling me what the interval is, then I will have a 95% probability that your interval contains the correct value. The key question is whether and how I should change this value when I learn which actual interval you obtained. It might be obvious that my probability should change - for example, the interval might be empty. If I understand the argument correctly, frequency Bayes analyses can sometimes be approximately preposterior valid while avoiding some of the nastier conditioning problems of this type. If the analyst has not done all of the heavy lifting which justifies serious informative conditioning, then this may serve as a "second best but much better than nothing" way of producing a somewhat informative probabilistic analysis.

[2] Yes, the subjectivist Bayesian approach provides a framework for a full consideration of all the relevant judgements for any uncertainty investigation, even where there is no conceivable frequentist criterion that could apply. Anyone who doubts this should look, for example, at Bayesian analysis for climate change and consider what frequentist interpretation these calculations could possibly have.

[3] Yes, default Bayesian analyses may approximate subjective inferences for large data sets or problems with genuinely weak prior information, although this must be established in each application.

[4] I haven't seen anything, in Jim's paper or elsewhere, which gives a non-subjective inferential meaning to default analysis which makes sense to me. It doesn't seem inconceivable that a different kind of meaning could be somehow developed though, for example viewing default Bayes as a useful exploratory data analysis tool.

[5] I guess this question does divide subjective and objective Bayesians. We would all be happy when we have data which overwhelms most reasonable priors and, as I indicated in my comments on Jim's paper, calling the conclusion of such an analysis "objective" doesn't seem outrageous. However, when the data is not so dominant, then there is no special status, in the subjective view, to the analysis with any particular default prior. This seems sensible to me. Reasonable people may disagree about the interpretation of the data and, if the answer is important, there is no privileged answer which can supersede the fully Bayesian analysis.

[6] No, for the reason spelt out by the next question. There is no other use of the term "objective" that is so logically compelling that its use wins out over the confusion that it causes.

[7] As long as we stick to the notion that the term "objective" refers to the strong property that everyone is compelled to agree, more or less, with the conclusions of the analysis, then this term may be helpful. When we allow "objective analyses" which disagree with the judgements of any informed person, then misunderstandings are inevitable.

[8] Yes. In decision-making we are concerned with making good decisions so that what matters is the results. In scientific inference, the reasoning which leads to the results is of crucial importance. In a decision problem, you could just accept that I was an expert, so that my decisions could automatically be acceptable to you. However, a scientist cannot simply claim to be correct. The difference is why a scientific Bayes analysis with limited data needs to be fully Bayesian, requiring not only a prior assessment, but also a careful justification of that assessment, and in addition an expression of the variation in belief across the scientific community.

[9] Of course, elicitation is invaluable, subject to the proviso that the aim is to provide not merely judgements but judgements whose rationale is made explicit. This is no more or less mysterious than the process of developing scientifically justified models. It's just less familiar.

5 Response to Larry Wasserman

Larry opens his comments with a fragment of a play. The work is not attributed, but I detect the hand of Samuel Beckett. As with all of Beckett's work, the meaning is not immediately obvious, and we need to read the text closely to make sense of the interchange between the protagonists. In as much as I can understand this cryptic encounter, it appears that several thousand 95% intervals were evaluated for physical quantities. (The play does not make this clear, but in the commentary in the next section Larry appears to indicate that these were unrelated problems.) For no apparent reason, it turns out that almost all of these intervals were incorrect. In immediate succession, this strange occurrence impoverishes the scientist and then, for mysterious reasons, it instead impoverishes the statistician. Much alcohol is consumed.

What is the meaning of this curious drama? Why were so many intervals wrong?

Was the scientist stupid? Was the statistician careless and not fully Bayesian? Was the current state of theory in physics wrong in so many thousands of fundamental aspects, in ways which, curiously, were enormously relevant to prior judgements for parameter values but which appear to have left the model likelihoods unchanged? Or is this drama meant to express the view that reality is so capricious that however carefully we use our expert judgements, exploiting all that we know from experiment, theory and common sense, the mere attempt to inject our understanding into the analysis is so doomed to failure and eventual disgrace that we would do better to abandon all pretence to know anything at all, and to declare support blindly for whatever the data appear to reveal, however contrary to our best current judgements this appears to be?

While it is an interesting theoretical exercise to imagine what would be the most rational form of inference in a truly "Beckettian" universe (and full credit to Larry for widening the discussion in this way!), I was discussing the best approach to inference in the rather more prosaic world that we actually inhabit. In this world, experts certainly do have relevant "subjective knowledge" which we may potentially exploit to make better inferences. There are many serious practical issues in the careful and appropriate use of such information — which is why to be a good Bayesian statistician requires an extra collection of skills as compared to a more traditional statistician — but the potential advantages from such effort are enormous.

Therefore, to address Larry's thought experiment, yes, I would prefer a sequence of fully Bayesian subjective analyses to a series of analyses with good frequency coverage that ignore all subjective input. I don't consider that the subjectivist analysis should have poor long run performance. For example, a priori, I typically expect my intervals to be shorter and better centered than frequency based alternatives, as I have tuned my procedures to behave well over the range of possible parameter values of greatest scientific interest. As I have corrected for obvious data irregularity, and used relevant historical information, a simple coverage count of whether the interval covers the true value or not is an oversimplification of the performance criteria that I would demand from the inference. We should criticise the non-subjective approach in wasting part of the data in simply re-establishing what we have already discovered so that we require much more data in order to be able to explore the likelihood surface in the actual region of scientific interest. This is separate from the diagnostic issue that we should carefully monitor whether we are actually achieving such improvements over non-subjective alternatives in practice as this will be informative for the validity of our wider state of knowledge.

So, to be clear, my fundamental objection to frequency Bayes is that, if I have carefully carried out a gold standard fully Bayesian subjective analysis, and this disagrees with the results of a frequency based analysis which ignores all of my subjective knowledge, then it would be bizarre for me to prefer the frequency answer. I'm sorry but you will have to work much harder to convince me that you have the power of ESP than to convince me that you have the power to distinguish Brahms from Elgar, and if your methods don't build this requirement in, then I don't like or trust them. In the meeting, it was argued that in many areas, such as drug regulatory authorities, it was precisely long run performance that was of concern. I doubt that this is true. Suppose

that a very careful fully Bayesian analysis suggested that a particular drug was potentially very dangerous, because of the mass of prior information suggesting bad effects for very similar drugs, but that the rather lazier regime of the frequency test, which could simply ignore all of this evidence, passed the drug. I doubt that the regulatory authority, having made the effort to carry out the fully Bayesian analysis, would be happy to license the drug on the argument that this likely mistake would be balanced out by some other judgements on unrelated drugs. An interesting field for lawyers!

Finally Larry asks about my view on elicitation. Elicitation has a central role in subjective Bayesian analysis. My only concern comes with those methods which treat prior probability distributions like pieces of data which can be measured simply by asking the relevant experts. We do not usually elicit likelihood functions in this way - rather we engage in careful scientific modelling, going through several iterations, seeking careful justifications for our choices, considering the implications of different choices and applying careful diagnostic scrutiny. Similarly, if there is substantial scientific experience on which the prior distribution should rest, then we need to carry out a careful modelling exercise to make this support explicit, both to reach judgements which are indeed justified by the current state of knowledge and also to make explicit and convincing to others the choice of prior judgements. In other words, elicitation is part of the usual hard work of careful science, and good elicitation methods, in my view, are those which meet such requirements in a clear and compelling fashion.

6 Response to comments by Fienberg, Kadane, Lad and O'Hagan

I have grouped the remaining discussants together simply because I have nothing further to add to their elegant and convincing support for the subjectivist position and their scepticism as to the status of the objectivist position as a helpful general construct.