Subjective Bayesian Analysis: Principles and Practice

Michael Goldstein
Department of Mathematical Sciences
University of Durham, UK

Abstract

We address the position of subjectivism within Bayesian statistics. We argue, firstly, that the subjectivist Bayes approach is the only feasible method for tackling many important practical problems. Secondly, we describe the essential role of the subjectivist approach in scientific analysis. Finally, we consider possible modifications to the Bayesian approach from a subjectivist viewpoint.

Keywords coherency, exchangeability, physical model analysis, high reliability testing, objective Bayes, temporal sure preference

1 Introduction

The subjective Bayesian approach is based on a very simple collection of ideas. You are uncertain about many things in the world. You can quantify your uncertainties as probabilities, for the quantities you are interested in, and conditional probabilities for observations you might make given the things you are interested in. When data arrives, Bayes theorem tells you how to move from your prior probabilities to new conditional probabilities for the quantities of interest. If you need to make decisions, then you may also specify a utility function, given which your preferred decision is that which maximises expected utility with respect to your conditional probability distribution.

There are many compelling accounts explaining how and why this view should form the basis for statistical methodology; see, for example, [11] and the accompanying discussion. In particular, [10] provides an excellent introduction to the subjectivist viewpoint, with a wide ranging collection of references to the development of this position. There is also a rich literature on elicitation, dealing with how generalised expert knowledge may be converted into probabilistic form; for a recent overview, see [4].

Moving from principles to practice, however, can prove very challenging and so there are many flavours of Bayesianism reflecting the technical challenges and requirements of different fields. In particular, a form of Bayesian statistics, termed "objective Bayes" aims to gain the formal advantages arising from the structural clarity of the Bayesian approach without paying the "price" of introducing subjectivity into statistical analysis. Such attempts raise important questions as to the role of subjectivism in Bayesian statistics. This account is my subjective take on the issue of subjectivism.

My treatment is split into three parts. Firstly, the subjectivist Bayes approach is the only feasible method for tackling many important practical problems, and in section 2 I'll give examples to illustrate this. Next, in section 3, I'll look at scientific analyses, where the role of subjectivity is more controversial, and argue the necessity of the subjective formulation in this context. In section 4, I'll consider how well the Bayes approach stands up to criticisms from the subjective viewpoint itself. Finally, in section 5, I'll comment on the implications of these developments.

2 Applied subjectivism

Among the most important growth areas for Bayesian methodology are those applications that are so complicated that there is no obvious way to even formulate a more traditional analysis. Such applications are widespread; for many examples, consult the series of Bayesian case studies volumes from the Carnegie Mellon conference series. Here are just a couple of areas that I have been personally involved in, with colleagues at Durham, chosen so that I can discuss, from the inside, the central role played by subjectivity.

2.1 High reliability testing for complex systems

Suppose that we want to test some very complicated system - a large software system would be a good example of this. Software testing is a crucial component of the software creation cycle, employing large numbers of people and consuming much of the software budget. However, while there is a great deal of practical expertise in the software testing community, there is little rigorous guidance for the basic questions of software testing, namely how much testing does a system need, and how can we design an efficient test suite for this purpose. The number of tests that we could, in principle, carry out is enormous, each test has non-trivial costs, both in time and money, and we must plan testing (and retesting given each fault we uncover) to a tight time/money budget. How can we design and analyse an optimal test suite for the system?

This is an obvious example of a Bayesian application waiting to happen. There is enormous uncertainty and we are forced to extrapolate beliefs about the results of all the tests that we have not carried out from the outcomes of the relatively small number of tests that we do carry out. There is a considerable amount of prior knowledge carried by the testers who are familiar with the ways in which this software is likely to fail, both from general considerations and from testing and field reports for earlier generations of the software. The expertise of the testers therefore lies in the informed nature of the prior beliefs that they hold. However, this expertise does not extend to an ability to analyse, without any formal support tools, the conditional effect of test observations on their prior beliefs, still less to an ability to design a test system to extract optimum information from this extremely complex and interconnected probabilistic system.

A Bayesian approach proceeds as follows. Firstly, we construct a Bayesian belief net. In this net the ancestor nodes represent the various general reasons that the testers may attribute to software failure, for example incorrectly stripping leading zeroes from a number. The links between ancestor nodes show relationships between these types of failure. The child nodes are the various test types, where the structuring ensures that all tests represented by a given test node are regarded exchangeably by the testers. Secondly, we quantify beliefs as to the likely levels of failure of each type and the conditional effects of each failure type on each category of test outcome. Finally, we may choose a test suite to optimise any prespecified criterion, either based on the probability of any faults remaining in the system or on the utility of allowing certain types of failure to pass undetected at software release. This optimality calculation is tractable even for large systems. This is because what concerns us, for any test suite, is the probability of faults remaining given that all the chosen tests are successful, provided any faults that are detected will be fixed before release.

In principle, this methodology, by combining Bayesian belief networks with optimal experimental design, is massively more efficient and flexible than current approaches. Is the approach practical? From our experiences working with an industrial partner, BT, I would say definitely yes. A general overview of the approach that we developed is given in [14]. As an indication of the potential increase in efficiency, we found that Bayesian automatic design provided 8 tests which together were more efficient than 233 tests designed by the original testing team, and identified additional tests that were appropriate for checking areas of functionality that had not been covered by the original test suite. This is not a criticism of the testers, who were very experienced, but simply illustrates that optimal multi-factor probabilistic design is very difficult. Therefore, the right tools offer enormous potential for efficiency gains.

Of course, there are many issues that must be sorted out before such benefits can be realised, from the construction of user-friendly interfaces for building the models to (a much larger obstacle!) the culture change required to recognise and routinely exploit such methods. However, the subjective Bayes approach does provide a complete framework for quantifying and managing the uncertainties of high-reliability testing. It is hard to imagine any other approach which could do so.

2.2 Complex physical systems

Many large physical systems are studied through a combination of mathematical modelling, computer simulation and matching against past physical data, which can, hopefully, be used to extrapolate future system behaviour; for example, this accounts for much of what we claim to know about the nature and pace of global climate change. Such analysis is riddled with uncertainty. In climate modelling, each computer simulation can take between days and months, and requires many input parameters to be set, whose values are unknown. Therefore, we may view computer simulations with varied choices of input parameters as a small sample of evaluations from a very high dimensional unknown function. The only way to learn about the input parameters is by matching simulator output to historical data, which is, itself, observed with error. Finally, and often most important, the computer simulator is just a model, and we need to consider the ways in which the model and reality may differ.

Again, the subjectivist Bayesian approach offers a framework for specifying and synthesising all of the uncertainties in the problem. There is a wide literature on the probabilistic treatment of computer models; a good starting point with a wide collection of references is the recent volume [13]. Our experience at Durham started with work on oil reservoir simulators, which are constructed to help with all the problems involved in efficient management of reservoirs. Typically, these are very high dimensional computer models which are very slow to evaluate. The approach that we employed for reservoir uncertainty analysis was based on representing the reservoir simulator by an emulator. This is a probabilistic description of our beliefs about the value of the simulator at each input value. This is combined with statements of uncertainty about the input values, about the discrepancy between the model and the reservoir and about the measurement uncertainty associated with the historical data. This completely specified stochastic system provides a formal framework allowing us to synthesise expert elicitation, historical data and a careful choice of simulator runs. Modulo many challenging technical issues arising from the size and complexity of the system, this specification allows us to identify “correct” settings for simulator inputs (often termed history matching in the oil industry), see [1], and to assess uncertainty for forecasts of future behaviour of the physical system, see [2]. Our approach relies on a Bayes linear foundation (which I’ll discuss in section 4) to handle the technical difficulties involved with the high dimensional analysis; for a full Bayes approach for related problems, see [9].

Our approach has been implemented in software employed by several users in the oil industry, through our collaborators ESL (Energy SciTech Limited). This means that we get to keep track, just a little, of how the approach works in practice. Here's an example of the type of success which ESL has reported to us. They were asked to match an oil field containing 650 wells, based on 1 million plus grid cells (for each of which permeability, porosity, fault lines, etc. are unknown inputs). Finding the previous best history match had taken one man-year of effort. Our Bayesian approach, starting from scratch, found a match using 32 runs (each lasting 4 hours and automatically chosen by the software), with a fourfold improvement according to the oil company measure of match quality. This kind of performance is impressive, although, of course, these remain very hard problems and much must still be done to make the approach more flexible, tractable and reliable.

Applications such as these make it clear that careful representation of subjective beliefs can give much improved performance in tasks that people are already trying to do. There is an enormous territory where subjective Bayes methods are the only feasible way forward. This is not to discount the large amount of work that must often be done to bring an application into Bayes form, but simply to observe that for such applications there are no real alternatives. In such cases, the benefits from the Bayesian formulation are potentially very great and clearly demonstrable. The only remaining issue, therefore, is whether such benefits outweigh the efforts required to achieve them. This "pain to gain" ratio is crucial to the success of subjective Bayes applications. When the answer really matters, such as for global climate change, the pain threshold would have to be very high indeed to dissuade us from the analysis.

By explicitly introducing our uncertainty about the ways in which our models fall short of reality, the subjective Bayes analysis also does something new and important. Only technical experts are concerned with how climate models behave, while everybody has an interest in how global climate will actually change. For example, the Guardian newspaper leader on Burying Carbon (Feb 3, 2005) tell us that "the chances of the Gulf Stream - the Atlantic thermohaline circulation that keeps Britain warm - shutting down are now thought to be greater than 50% ". This sounds like something we should know. However, I am reasonably confident that no climate scientist has actually carried an uncertainty analysis which would be sufficient to provide a logical bedrock for such a statement. We can only use the analysis of a global climate model to guide rational policy towards climate change if we can construct a statement of uncertainty about the relation between analysis from the climate model and the behaviour of the real climate. To further complicate the assessment, there are many models for climate change in current use, all of whose analyses should be synthesised as the basis of any judgements about actual climate change. Specifying beliefs about the discrepancy between models and reality is unfamiliar and difficult. However, we cannot avoid this task if we want our statements to carry weight in the real world. A general framework for making such specifications is described in [8].

3 Scientific subjectivism

3.1 The role of subjectivism in scientific enquiry

In the kind of applications we've discussed so far, the only serious issues about the role of subjectivity are pragmatic ones. Each aspect of the specification, whether part of the "likelihood function" or the "prior distribution", encodes a collection of subjective judgements. The value of the Bayesian approach lies firstly in providing a language within which we can express all these judgements and secondly in providing a calculus for analysing these judgements.

Controversy over the role of subjectivity tends to occur in those areas of scientific experimentation where we do appear to have a greater choice of statistical approaches. Laying aside the obvious consideration that any choice of analysis is the result of a host of subjective choices, there are, essentially, two types of objections to the explicit use of subjective judgements; those of principle, namely that subjective judgements have no place in scientific analyses; and those of practice, namely that the pain to gain ratio is just too high.

These are deep issues which have received much attention; a good starting place for discussion of the role of Bayesian analysis in traditional science is [12]. Much of the argument can be captured in simple examples. Here's one such, versions of which are often used to introduce the Bayesian idea to people who already have some familiarity with traditional statistical analysis.

Firstly, we can imagine carrying out Fisher's famous tea-tasting experiment. Here an individual, Joan say, claims to be able to tell whether the milk or the tea has been added first in a cup of tea. We perform the experiment of preparing ten cups of tea, choosing each time on a coin flip whether to add the milk or tea first. Joan then tastes each cup and gives an opinion as to which ingredient was added first. We count, the number, X, of correct assessments. Suppose, for example, that $X = 9$.

Now compare the tea tasting experiment to an experiment where an individual, Harry say, claims to have ESP as demonstrated by being able to forecast the outcome of fair coin flips. We test Harry by getting forecasts for ten flips. Let X be the number of correct forecasts. Suppose that, again, $X = 9$.

Within the traditional view of statistics, we might accept the same formalism for the two experiments, namely that, for each experiment, each assessment is independent with probability p of success. In each case, X has a binomial distribution parameters 10 and p, where $p = 1/2$ corresponds to pure guessing. Within the traditional approach, the likelihood is the same, the point null is the same if we carry out a test for whether $p = 1/2$, and confidence intervals for p will be the same.

However, even without carrying out formal calculations, I would be fairly convinced of Joan's tea tasting powers while remaining unconvinced that Harry has ESP. You might decide differently, but that is because you might make different prior judgements. This is what the Bayesian approach adds. Firstly, we require our prior probability, g say, that Harry or Joan is guessing. Then, if not guessing, we need to specify a prior distribution q over possible values of p. Given g, q, we can use Bayes theorem to update our probability that Harry or Joan is just guessing and, if not guessing, we can update our prior distribution over p. We may further clarify the Bayesian account by giving a more careful description of our uncertainty within each experiment based on our judgements of exchangeability for the individual outcomes. This allows us to replace our judgements about the abstract model parameter p with judgements about observable experimental outcomes as the basis for the analysis.

Therefore, the Bayes approach shows us exactly how and where to input our prior judgements. We have moved away from a traditional view of a statistical analysis, which attempts to express what we may learn about some aspect of reality by analysing an individual data set. Instead, the Bayesian analysis expresses our current state of belief based on combining information from the data in question with whatever other knowledge we consider relevant.

The ESP experiment is particularly revealing for this discussion. I used to use it routinely for teaching purposes, considering that it was sufficiently unlikely that Harry would actually possess ESP that the comparison with the tea tasting experiment would be self-evident. I eventually came to realise that some of my students considered it perfectly reasonable that Harry might possess such powers. While writing this article, I tried googling "belief in

ESP" over the net, which makes for some intriguing reading. Here's a particularly relevant discussion from an article in the September 2002 issue of Scientific American, by Michael Sherme, titled "Smart People Believe Weird Things". After noting that, for example, around 60% of college graduates appear to believe in ESP, Sherme reports the results of a study that found "no correlation between science knowledge (facts about the world) and paranormal beliefs. The authors, W. Richard Walker, Steven J. Hoekstra and Rodney J. Vogl, concluded: "Students that scored well on these [science knowledge] tests were no more or less sceptical of pseudo-scientific claims than students that scored very poorly. Apparently, the students were not able to apply their scientific knowledge to evaluate these pseudo-scientific claims. We suggest that this inability stems in part from the way that science is traditionally presented to students: Students are taught what to think but not how to think." Sherme continues as follows. "To attenuate these paranormal belief statistics, we need to teach that science is not a database of unconnected factoids but a set of methods designed to describe and interpret phenomena, past or present, aimed at building a testable body of knowledge open to rejection or confirmation."

The subjective Bayesian approach may be viewed as a formal method for connecting experimental factoids. Rather than treating each data set as though it has no wider context, and carrying out each statistical analysis just as though this were the first investigation that had ever been carried out of any relevance to the questions at issue, we consider instead how the data in question adds to, or changes, our beliefs about these questions.

If we think about the ESP experiment in this way, then we should expand the problem description to reflect this requirement. Here is a minimum that I should consider. First, I would need to assess my probability for E, the event that ESP is a real phenomenon that at least some people possess. This is the event that joins my analysis of Harry's performance with my generalised knowledge of the scientific phenomenon at issue. Conditional on E, I should evaluate my probability for J, the event that Harry possesses ESP. Conditional on J and on J complement, I should evaluate my probabilities for G, the event that Harry is just guessing and C, the event that either the experiment is flawed or Harry is, somehow, cheating; for example, the coin might be heads biased and Harry mostly calls heads. This is the event that captures my generalised knowledge of the reliability of experimental procedures in this area. If there is either cheating or ESP, I need a probability distribution over the magnitude of the effect.

What do we achieve by this formalism? Firstly, this gives me a way of assessing my actual posterior probability for whether Harry has ESP. Secondly, if I can lay out the considerations that I use in a transparent way, it is easy for you to see how your conclusions might differ from mine. If we disagree as to whether Harry has ESP, then we can trace this disagreement back to differing probabilities for the general phenomenon, in this case ESP, or different judgements about particulars of the experiment, such as Harry's possible ability at sleight of hand. More generally, by considering the range of prior judgements that might reasonably be made, I can distinguish between the extent to which the experiment might convince me as to Harry's ESP, and the effect it might have on others. I could even determine how large and how stringently controlled an experiment would need to be in order to have a chance of convincing me of Harry's powers. More generally, how large would the experiment need to be to convince the wider community?

The above example provides a simple version of a general template for any scientific Bayesian analysis. There are scientific questions at issue. Beliefs about these issues require prior specification. Then we must consider the relevance of the scientific formulation to the current experiment along with all the possible flaws in the experiment which would invalidate the analysis. Finally, a likelihood must be specified, expressing data variability given the hypotheses of interest. There are two versions of the subsequent analysis. Firstly, you may only want to know how to revise your own beliefs given the data. Such private analyses are quite common. Many scientists carry out at least a rough Bayes assessment of their results, even if they never make such analysis public. Secondly, you may wish to publish your results, to contribute to, or even to settle, a scientific issue. It may be that you can construct a prior specification that is very widely supported. Alternately, it may be that, as with the ESP experiment, no such generally agreed prior specification may be made. Indeed, the disagreement between experts may be precisely what the experiment is attempting to resolve. Therefore, our Bayesian analysis of an experiment should begin with a probabilistic description whose qualitative form can be agreed on by everyone. This means that all features, in the prior and the likelihood, that cause substantial disagreement should have explicit form in the representation, so that differing judgements can be expressed over them. Statistical aspects of the representation may employ standard data sharing methodologies such as meta-analysis, multi-level modelling and Bayes empirical Bayes, provided all the relevant judgements are well sourced. We can then produce the range of posterior judgements, given the data, which correspond to the range of "reasonable" prior judgements held within the scientific community. We may argue that a scientific case is "proven" if the evidence should be convincing given any reasonable assignment of prior beliefs. Otherwise, we can assess the extent to which the community might still differ given the evidence. We should make this analysis at the planning stage in order to design experiments that can be decisive for the scientific community or to conclude that no such experiments are feasible.

All of this is clear in principle, though implementation of the program may be difficult in individual cases. Each uncertainty statement is a well sourced statement of belief by an individual. If individual judgements differ and if this is relevant, then such differences are reflected in the analysis. However, in practice such an approach is rarely followed. Let us consider again the objections to the Bayesian approach.

3.2 Potential objections to scientific subjectivism

The principled objection to Bayesian subjectivism is that the subjective Bayesian approach answers problems wrongly, because of unnecessary and unhelpful appeals to arbitrary prior assumptions, which should have no place in scientific analyses. Individual subjective reasoning is inappropriate for reaching objective scientific conclusions, which form the basis of consensus within the scientific community.

This objection would perhaps have more force if there was a logically acceptable alternative. I do not here want to dwell on the difficulties in interpretation of the core concepts of more traditional inference, such as significance and coverage properties: a valid confidence interval may be empty, for example when constructed by the intersection of a series of repeated confidence intervals; a statistically significant result obtained with high power may be almost certainly false, and so forth. Further, I do not know of any way to construct even the basic building blocks of the inference, such as the relative frequency probabilities that we must use if we reject the subjective interpretation, that will stand up to proper logical scrutiny. Instead, let us address the principled objection directly. We cannot consider whether the Bayes approach is appropriate without first clarifying the objectives of the analysis. When we discussed the analysis of physical models, we made the fundamental distinction between analysis of the model and analysis of the physical system. Analysing various models may give us insights but at some point these insights must be integrated into statements of uncertainty about the system itself. Analysing experimental data is essentially the same. We must be clear as to whether we are analysing the experiment or the problem.

In the ESP experiment, the question is whether Harry has ESP, or, possibly, whether ESP exists at all. If we analyse the experimental data as part of a wider effort to address our uncertainty about these questions, then external judgements are clearly relevant. As described above, the beliefs that are analysed may be those of an individual, if that individual can make a compelling argument for the rationality of a particular belief specification, or instead we may analyse the collection of beliefs held by informed individuals in the community. The Bayes analysis is appropriate for this task, as it is concerned to evaluate the relevant kinds of uncertainty judgements, namely the uncertainties over the quantities that we want to learn about, given the quantities that we observe, based on careful foundational arguments using ideas such as coherence and exchangeability to show why this is the unavoidable way to analyse our actual uncertainties.

On the other hand, suppose that, for now, we only want to analyse the data from this individual experiment. Our goal, therefore, cannot be to consider directly the basic question about the existence of ESP. Indeed, it is hard to say exactly what our goal is, which is why there often is so much confusion in discussions between proponents of different approaches. All that we can say informally is that the purpose of such analysis is to provide information which will be helpful at some future time for whoever does attempt to address the real questions of interest. We are now in the same position as the modeller; we have great freedom in carrying out our analyses but we must be modest in the claims that we make for them.

This is the world in which we find objective Bayes methodology. What does the word "objective" mean in this context? It does not mean that there is an objective status for the statements made by the methodology, as the approach doesn't offer any other testable meaning for probability statements beyond the uncertainty judgements of the individual. Nor does it mean that there is some objectively testable property that the answers derived by the analysis will necessarily satisfy. Thus, we have no way to judge in what sense, and to what degree, we should have confidence in the conclusions of an objective Bayes analysis. It does not even mean that there is some objectively testable principle that has been used to assemble the ingredients of the analysis.

Instead, as with most other uses of the term, objective here usually means that we are not attempting to address the question at issue (should we think that Harry has ESP?) but instead we are constructing a model for the inference by introducing and attempting to answer some surrogate question which is less challenging. I'm not sure what the question would be here - perhaps we imagine somebody who has just arrived on this planet and we wonder what our stranger would conclude if immediately confronted by Harry's performance. Of course, if we formulate this surrogate question too precisely then we will not be able to answer it; after all, we have no idea what such a stranger would actually conclude. This ambiguity can sometimes be benign. If we have a very large experiment, then a simple automatic choice of prior, along with some large sample approximation argument to show that the conclusion is not overly sensitive to the choice of prior, may save a lot of time and effort as compared to a full subjectivist analysis while reaching substantially the same conclusion. However, the value of such an analysis still lies in the robust approximation to the full subjectivist analysis.

When analysis of the current experiment is sensitive to the prior specification, as in our ESP experiment, it is clear that there is no objective answer to the question of Harry's powers, based on analysis of the given data. To pretend otherwise is to enter the world of pseudo-science which we alluded to above, in which we behave just as those science students who appear unable to make the links between reason, experience and observation. Subjective Bayesian analysis is hard but necessary precisely because it does concern such a fully rounded assessment.

Of course, the practical objection to routine use of subjective Bayesian analysis is that it is too hard, because of the difficulty of finding justifiable prior distributions for the quantities of interest in complicated problems. If the questions at issue are not sufficiently important to warrant a full subjectivist analysis, or if a simple analysis can bring out the most important messages of the data very quickly, or if the area is sufficiently new that our views really are like those of a stranger from another planet, then some form of automatic Bayes analysis may be useful and revealing. However, just as with over-reliance on any other model, the danger in relying too much on such automatic analyses is that we will forget their limitations.

And if statisticians risk confusion as to the meaning of their analyses, how much greater is the danger for those non statisticians who rely on the output of statistical analyses? As a current and tragic example, the General Medical Council for the UK has just ruled that Professor Sir Roy Meadow should be struck off the medical register due to serious professional misconduct for giving evidence beyond his expertise at a trial which led to a mother being wrongfully jailed for the murder of her two baby sons. Much of the evidence of misconduct is based around Prof Meadow's statement at the trial that there was just a "one in 73 million" chance that two babies with the given background could each suffer cot death. (This figure was apparently obtained by squaring the circa 8,500 to one chance of a single baby dying of cot death in a family.) The actual odds are now considered to be far lower. The GMC fitness to practise panel said in its verdict that Prof Meadow had failed in his duty to check the validity of his statistics. There are many features of the case which are worthy of comment. Of particular relevance is Prof. Meadows defence of his claim. The following comes from the Guardian, July 2nd, 2005.

"Prof Meadow, whose evidence was used in the cases of three other women wrongly accused of killing their babies, said he had been quoting the statistic from a highly respected report on sudden infant deaths, which at the time had yet to be published. Defending his right to use the report in his evidence at Mrs Clark's trial, he said, "I was quoting what I believed to be a very thorough study ... by experts, several of whom I knew and respected." Nicola Davies QC, representing Prof Meadow, asked: "Did you have any difficulty with quoting statistics from the study?" He replied: "To me it was like I was quoting a radiologist's report or a piece of pathology ... I was quoting the statistics, I wasn't pretending to be a statistician.""

I have not seen the study in question, although I have read claims that Prof Meadow quoted some calculations from the study which were taken out of context, ignoring the conditions and qualifications around the quoted values. However, the general attitude displayed to the statistical analysis, conferring on it a purely objective and value free status, surely lies at the heart of the issue. Prof Meadow's professional misfortune may only have been that the statistical 'mistake' for which he is blamed was sufficiently elementary that it could easily be argued that overlooking the error was professionally negligent. I can easily envisage a more sophisticated treatment, say an 'objective Bayes' analysis, which, by placing 'uninformative' priors on certain key parameters in a more elaborate version of the model, could make essentially the same 'error' but in a way which would be far harder to detect. As statistical analyses become more sophisticated and more difficult for anyone but an expert statistician to check, it becomes increasingly important that the meaning of the statistical analysis is clearly conveyed. Any statistician who does a Bayesian analysis for a problem with important practical consequences but does not make good and clear use of informed judgements, and then labels that analysis as 'objective', should be aware of the misunderstandings and mistakes that will follow when their claim is taken precisely at face value.

4 Pure subjectivism

We have argued that the subjective Bayes approach is successful in practice, and is invaluable for serious scientific analysis. This, however, leaves open possible criticisms of the Bayes approach from the subjective viewpoint itself. Carrying out a careful Bayesian analysis can prove very difficult. In part this is because such an analysis requires an extremely detailed level of prior probabilistic belief specification. Typically, we find it difficult to make detailed specifications in a way which genuinely corresponds to our prior beliefs. Therefore, artificial, conventional prior forms are used, often bearing only a limited relation to our prior beliefs, so that the resulting Bayesian analysis bears only a limited relation to our actual judgements.

Is such a detailed specification really necessary? A true subjectivist formulation might start by recognising the limited abilities of the individual to make large collections of uncertainty specifications. It is precisely this consideration that led de Finetti, in [3], the most careful statement of the subjectivist position yet written, to chose expectation rather than probability as his primitive for the subjectivist theory. With expectation as primitive, we can assess directly whatever sub-collection of probabilities and expectations we consider ourselves able to specify at any given time, whereas, if probability is the primitive, then we must specify every probability before we can specify any expectation. Unfortunately, the liberating aspect of this approach is somewhat lost in de Finetti's development, as changes in beliefs are still carried out by conditioning on events, which again requires a finely detailed level of prior specification (although we may give bounds on the coherent inferences consistent with any partial specification, see [10]).

Therefore, we must also consider whether it is an intrinsic part of the subjectivist position that beliefs should change by conditioning via Bayes theorem. This question cuts to the heart of the Bayes position, as it is impossible to demonstrate any definitive sense in which beliefs should change by conditioning. You might think that someone, somewhere has proved that conditioning is the correct way to modify beliefs, at least under certain conditions. However, all that can be proved is results such as the following. Suppose that you specify a joint probability distribution for a collection of random quantities. Suppose that you also write down a rule for changing your probabilities for some of the quantities, as a function of the numerical values of the remaining quantities. If this rule for changing your probabilities is not the usual conditional probability formula, then you can be made a sure loser, in the usual sense of placing a sequence of bets that pay off depending on various combinations of outcomes of the random quantities.

This is not a demonstration that beliefs should change by conditioning: all that it does is to eliminate non-Bayesian rules for updating beliefs in the class of rules based exclusively on current beliefs and the values of the observables. The fundamental question remains as to what relevance probabilities that are declared conditional on the outcome of certain events should hold for the actual posterior probabilities that you assign when you do learn of the outcomes. By the time that you observe the data, you may have come across further unanticipated but relevant information (or you may not, and this also is relevant information), and you may well have further general insights about the problem, by study of relevant literature, deeper mathematical treatment or careful data analysis. None of this corresponds to Bayesian conditioning. Indeed, I cannot remember ever seeing a non-trivial Bayesian analysis which actually proceeded according to the usual Bayes formalism. All of this illustrates the simple observation that there is no stronger reason why there should be a rule for going from prior to posterior beliefs than that there should be such a rule for constructing prior beliefs in the first place. (For example, any attempt to view conditional beliefs as the beliefs that you "should" hold were you to observe the conditioning event and "nothing else" is doomed to self-contradiction, as the fact that you observed nothing else was not part of the original conditioning event, and would be informative were it to be included in the conditioning.)

For the above reasons, all attempts to present the Bayesian approach as a normative theory, which describes how we should, in principle, modify our beliefs given evidence, must be fundamentally incomplete. They are analogous to a similar discussion as to whether and when, say, a global climate model is right or wrong. This is the wrong question. We know that the global climate model differs from the actual climate - they are two quite different things. Instead, our two tasks are firstly to identify why we consider that a particular climate model is informative for climate, and secondly to quantify the value of this information, by considering the residual uncertainty in climate behaviour, given the analysis of the climate model.

If we view the Bayes formalism as providing a model for belief change which is neither normative nor descriptive, then the natural questions are firstly why do we consider this model relevant to actual problems of belief change and secondly how do we describe the discrepancy between this model and the reality of changing beliefs? Our answers to these questions are as follows; for details see [6]. We begin by distinguishing between your current conditional probabilities $\mathrm{P}(A|B)$, $\mathrm{P}(A|B^c)$ and your posterior probability $\mathrm{P}_t(A)$ that you will assign at future time t after you have observed either B or B^c. We need a temporal principle to link beliefs that you specify now with those that you will specify at time t. This link is provided by the temporal sure preference (TSP) principle, which is as follows.

"If you are sure that at future time t you will prefer the (small) random penalty A to the (small) random penalty B, then you should not now prefer B to A."

TSP places a very weak requirement on your temporal preferences, which would certainly be satisfied within the conventional Bayesian formulation. However, unlike the Bayes formalism, which seeks to make today's specification logically compelling for tomorrow's revision of belief, TSP places our preferences in the right order, requiring logical certainty in the future to be compelling for our current belief evaluations.

If we accept TSP for the current inference, then we can show, see [6], that this establishes the following stochastic relationship between conditional and posterior probabilities, namely that at the present moment you must make the specification

$$\mathrm{P}_t(A) = \mathrm{P}(A|\mathcal{B}) + \mathcal{R} \tag{1}$$

where $\mathrm{P}(A|\mathcal{B})$ is the conditioning of A on the partition $\mathcal{B} = (B, B^c)$, namely

$$\mathrm{P}(A|\mathcal{B}) = \mathrm{P}(A|B)B + \mathrm{P}(A|B^c)B^c$$

where B, B^c are the indicator functions for the corresponding events, and R is a further random quantity with

$$\mathrm{E}(R) = \mathrm{E}(R|B) = \mathrm{E}(R|B^c) = 0$$

This corresponds closely to the interpretation that we have earlier suggested for mathematical models of physical systems. For example, a climate model does not tell us what will actually happen, but instead is useful, for example, in giving us a mean forecast, with associated variance, whose value lies in reducing, but not eliminating, our uncertainty about climate behaviour. Similarly, from (1), we are justified in viewing $\mathrm{P}(A|\mathcal{B})$ as providing a mean forecast for our future judgements, while the residual quantity R expresses the uncertainty in the conditional mean forecast. Informally, the larger the variance of $\mathrm{P}(A|\mathcal{B})$ as compared to the variance of R, the more informative a formal Bayes inference based on conditioning on B will be for the actual posterior judgement on A. As the variance of $\mathrm{P}_t(A)$ is fixed, we may both increase the variance of $\mathrm{P}(A|\mathcal{B})$ and decrease the variance of R by refining the partition $\mathcal{B}$.

This argument clarifies the logical status of a Bayesian analysis. It also frees us from the tyranny of conditioning. Even though (1) concerns probabilities, this relation can only be derived within a formalism which starts with expectation as primitive. In that formulation, probabilities are just expectation statements, and (1) is a special case of the following general result. Suppose that D is a vector of quantities which will be observed by time t. Given TSP, your current beliefs about your posterior expectation $\mathrm{E}_t(X)$ for any other random vector X, specified at time t, must satisfy the following relations:

$$X = \mathrm{E}_t(X) + S \tag{2}$$

$$\mathrm{E}_t(X) = \mathrm{E}_D(X) + R \tag{3}$$

where $\mathrm{E}_D(X)$ is the Bayes linear mean for X determined by

$$\mathrm{E}_D(X) = \mathrm{E}(X) + \mathrm{Cov}(X, D)(\mathrm{Var}(D))^{-1}(D - \mathrm{E}(D)$$

and R, S are further random quantities, with

$$\mathrm{E}(R) = \mathrm{E}(S) = \mathrm{Cov}(R, D) = \mathrm{Cov}(S, D) = \mathrm{Cov}(R, S) = \mathrm{Cov}(S, \mathrm{E}_t(X)) = 0$$

The Bayes linear analysis is based on direct specification of means, variances and covariances; for an overview of the Bayes linear approach to statistics see [7]. From (2), (3), the Bayes linear analysis for X bears the same relation to the actual posterior judgement for X that the posterior judgement for X bears to the quantity X itself. Bayesian conditioning is simply the special case of (3) in which D comprises the indicator functions for a partition.

If conditioning is not the operation underpinning the Bayesian analysis, then the requirement of full probabilistic specification can be seen as an arbitrary imposition. We may make full probabilistic specifications where this is natural and straightforward, and, this representation will maximise the proportion of uncertainty expressed by $\mathrm{E}_D(X)$. However, this is a refinement of degree, not of kind. If the extra information is worth the effort in prior specification and analysis that is required, then the full Bayes approach is worthwhile. Otherwise a simpler analysis is appropriate. Placing the subjectivist analysis within a logical framework which distinguishes between the model for the inference and the actual inference gives us control of the level of detail of our prior specification and analysis, while reminding us of the requirement to relate the formal analysis to the larger inferential problem which should always be our primary concern. Of course, this raises further questions as to precisely how such inferences should be conducted. This is largely unexplored territory; for theoretical underpinnings embedding statistical models derived from exchangeability judgements within this more fully subjectivist view, see [5]. Subjectivist theory offers a language and framework rather than a complete description of belief representation and inference. Whether such a complete description could ever be provided is, in my subjective opinion, extremely doubtful.

5 Concluding comments

The subjective Bayes approach is alive and well and proving very successful in many important practical applications. However, much of the potential of the approach is still to be realised. Subjectivist analysis may appear daunting, but, of course, what is difficult is making reasoned judgements about complex situations within any framework at all. The subjectivist approach does not make these difficulties vanish, but it does offer a coherent language and tool set for analysing all of the uncertainties in complicated problems, and therefore provides the best method that I know for analysing uncertainty in important real world problems. But who will carry out such analyses? Modellers are skilled at modelling, theorists develop theory, experimenters spend their time experimenting and statisticians tend to view their role as analysing data. These are all essential skills. But we are missing the specialism which moves beyond these comfort zones and puts all these activities together.

In section 2, we emphasised the distinction between analysing models and making inferences about the systems that the models purport to represent. In section 3, we made the similar distinction between making such inferences and analysing data collections of some relevance to these questions. In section 4, we further distinguished between our actual inferences about such systems and the output of formal inferential mechanisms, such as traditional Bayes. When we properly recognise, develop and apply the ideas and methods of subjectivist analysis, then we will be able to carry out that synthesis of models, theory, experiments and data analysis which is necessary to make real inferences about the real world.

References

[1] Craig, P.S., Goldstein, M., Smith, J.A. and Seheult, A. H.(1997)Pressure matching for hydrocarbon reservoirs: a case study in the use of Bayes linear strategies for large computer experiments (with discussion), in Case Studies in Bayesian Statistics, vol. III, eds. Gastonis, C. et al. 37-93. Springer-Verlag.

[2] Craig, P.S., Goldstein, M., Rougier, J. C. and Seheult, A. H. (2001) Bayesian Forecasting Using Large Computer Models, JASA, 96, 717-729

[3] de Finetti, B. (1974,1975) Theory of Probability, vols 1 & 2, Wiley

[4] P.H.Garthwaite, P.H. and J.B.Kadane, J.B. and O'Hagan. A., (2005) Elicitation, J. Amer. Statist. Assoc., to appear.

[5] Goldstein, M.(1994) Revising Exchangeable Beliefs: Subjectivist foundations for the inductive argument, in Aspects of uncertainty: a tribute to D.V. Lindley, 201-222, eds. Freeman, P.R. and Smith, A.F.M., Wiley, .

[6] Goldstein, M.(1997) Prior inferences for posterior judgements, in Structures and norms in Science, M.C.D. Chiara et. al. eds., 55-71, Kluwer.

[7] Goldstein, M. (1999) Bayes linear analysis , In Encyclopaedia of Statistical Sciences, update volume 3, 29-34, eds. Kotz, S., et al., Wiley.

[8] Goldstein, M. and Rougier, J. C. (2004) Probabilistic formulations for transferring inferences from mathematical models to physical systems SIAM journal on scientific computing, to appear.

[9] Kennedy, M. and O'Hagan, A. (2001). Bayesian calibration of computer models (with discussion). Journal of the Royal Statistical Society, Series B. 63, 425-464.

[10] Lad, F. (1996) Operational subjective statistical methods: a mathematical, philosophical and historical introduction, Wiley, New York

[11] Lindley, D.V. (2000) The philosophy of statistics (with discussion), The Statistician, 49,293-337

[12] Howson, C. and Urbach, P. (1989) Scientific reasoning: the Bayesian approach, Open Court

[13] Santner, T. J., Williams, B. J., Notz, W. I. (2003), The Design and Analysis of Computer Experiments, Springer

[14] Wooff, D.A., Goldstein, M. and Coolen, F. (2002) Bayesian Graphical Models for Software Testing, IEEE Transactions on Software Engineering, 28, 510-525.