WHY BE A BAYESIAN?

Michael Goldstein
Dept. of Mathematical Sciences, University of Durham,England

Abstract
This paper presents a discussion of the value of the Bayesian approach for scientific enquiry, through simple examples, general principles, and an overview of ideas which are useful for the Bayesian analysis of large physical models.

The Bayesian paradigm is, in many ways, the original approach to statistical reasoning. While, in the scientific arena, the approach has been viewed with suspicion, it is now spreading and finding acceptance in ever wider areas. We shall reappraise the strengths and weaknesses of the approach for scientific inference. Of course, this subject has a wide discussion literature, see, for example, [6]; for overviews of the Bayesian approach see [1], [8]. My intention here is to examine what I feel are the compelling reasons to use a Bayesian approach in relatively simple problems, and then to discuss what happens when such an approach becomes more difficult to follow.

1 THE BAYESIAN APPROACH

In the Bayesian approach, knowledge about unknown quantities of interest, θ, as expressed by a prior probability distribution $P(\theta)$, is combined with knowledge from data, d, expressed by a likelihood function $P(d|\theta)$, to give knowledge after seeing d expressed by a posterior distribution for θ evaluated as

$$P(\theta|d) \propto P(d|\theta)P(\theta) \tag{1}$$

There is no question as to the correctness of the theorem. However, the application provokes much controversy. Many of the arguments divide along the following lines: supporters of the approach argue that it is **Correct** and **Useful**, while opponents argue that it is **Inappropriate** and **Hard** to apply Bayes in the scientific arena. In summary, the arguments are:

BAYES IS CORRECT

[C1] Other approaches are wrong, as argued through the well-rehearsed counter-examples about the failure of meaning of the core concepts of more traditional inference, such as significance and coverage properties. Thus, a valid confidence interval may be empty, a statistically significant result obtained with high power may be almost certainly false, and so forth.

[C2] The Bayes approach is right, as argued on the grounds that the method evaluates 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 uncertainties

BAYES IS USEFUL

[U1] The methodology gives good solutions for standard problems, as argued through individual cases. The solutions appear paradox-free, and correspond well with intuition.

[U2] The methodology offers the only way to tackle many non-standard problems, as there is a unified approach for all problems in uncertainty. It offers a method which can always be followed, unlike most other approaches which rely on ad hoc tricks for each individual case.

BAYES IS INAPPROPRIATE

[I1] Bayesian methodology answers problems wrongly. Usually, this is attributed to unnecessary and unhelpful appeal to arbitrary prior assumptions, which should not belong in scientific analyses.

[I2] Bayesian methodology answers the wrong problems. This argument replaces the blanket criticism of the Bayes approach by recognition that the Bayes solution may indeed tell us something meaningful about what an individual might conclude from the data, but still argues that such individual subjective reasoning is inappropriate as a way of reaching sound and objective scientific conclusions, which are related to consensus within the scientific community.

BAYES IS HARD

[H1] Every problem is hard for Bayesian analysis. This is a reflection of the difficulty, even in the simplest problem, of finding an objectively justifiable prior distribution for the quantities of interest. In general how do we find prior distributions and what should we do if experts disagree?

[H2] Hard problems are hard for Bayesian analysis. Even if we could solve the prior specification issue for simple problems, the difficulty involved in constructing a full Bayes specification for more complicated problems renders the approach infeasible.

The above arguments have been simplified down to their essential form to suggest that there are (at least!) two levels at which we may debate the correct use of statistical methodology:

(i) the **current practice** debate: **[C1],[U1]**, versus **[I1],[H1]**

(ii)the **underlying issues** debate: **[C2],[U2]**,versus **[I2],[H2]**

Of course, the two debates are intimately linked, and starting in one debate we may easily find ourselves dipping into the other. However, unless we are clear as to which debate we are in, it is easy to become confused, especially as the structure of the two debates appears so similar.

The current practice debate is essentially pragmatic. We look at familiar problems, and from a common-sense practitioner's viewpoint try to evaluate the competing arguments. Proponents of the Bayesian argument find their solutions intuitively appealing, while being able to poke holes in solutions propounded by other methods, while opponents consider that the extra ingredients that Bayesians have introduced are at best irrelevant and arbitrary and at worst meaningless. This is a natural starting point for deciding which viewpoint to adopt. However, it is subsumed by the underlying issues debate, in the sense that if the approach really is correct and applies to a very wide range of problems, then of course we should use it, while if the approach focusses on the wrong problems and the wider range of applications is in practice infeasible, then we should be correspondingly sceptical. Each aspect of the debate has innumerable nuances. Here, I shall content myself with developing the arguments (not impartially - I am most definitely on the Bayes side) starting with a simple example relevant to the current practice debate, and then moving on to considering the underlying issues.

2 CURRENT PRACTICE: A DIAGNOSTIC EXAMPLE

Let us start with an example that is simple enough that we may all agree on the appropriate analysis. Suppose that you either have a particular (rare) disease (event D) or that you don't (event $\overline{D}$). You take a diagnostic test which gives either a positive response (event $+$), or a negative response (event $-$). The test is judged to be 99% reliable, e.g $P(+|D) = 0.99$, $P(+|\overline{D}) = 0.01$. You take the test and get a positive result. Do you have the disease? We shall develop the analysis through three stages, corresponding to the conceptual issues involved in applying the Bayesian argument.

[1]: Known disease rate. Suppose that it is known that the proportion, p, of people with the disease is 0.001. Then it is uncontroversial that we may apply Bayes theorem, giving

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|\overline{D})P(\overline{D})} = 0.09 \tag{2}$$

Everyone will agree that most likely you do not have the disease, but have scored a false positive.

[2]: Unknown disease rate. Now suppose that you don't know the disease proportion p. Suppose that you specify a prior probability distribution for p with expectation $\mathrm{E}(p) = 0.001$. As the value of

$\mathrm{E}(p)$ is all that is required for applying Bayes theorem, then you again have exactly the same numerical result, namely that you judge $P(D|+) = 0.09$. Does this result have the same meaning? For you, yes it does. You were uncertain beforehand as to your disease state, and the data has changed your uncertainty, in a well-defined way, and there is no operational difference for you between cases [1] and [2], unless you obtain further information relevant to the general prevalence of the disease. However, we have not specified exactly how you reached your prior assessment. Someone who disagreed with you about the rarity of the disease might come to a very different conclusion. Therefore, we might ask how high a belief in the disease, a priori, you would need to have at least a 50% probability of having the disease after seeing the positive test; from (2), you would need a prior expectation of $\mathrm{E}(p) > 0.01$. Similarly you would need to have at least a 50% probability of having the disease a priori to give a 99% probability of the disease given a positive test. Thus, there are now two aspects to the inference - how you, as an individual, should react to the data, and how a wider community might so react, which is addressed through a sensitivity analysis.

[3]: Pervasive disease. Now suppose that you know that either p is 0 or 1, that is everybody or nobody has the disease. This is just an special case of [2], and again all that matters is $\mathrm{E}(p) = P(p = 1)$, in this case. Otherwise, for you cases [2] and [3] are the same. However, case [3] leads directly into the scientific version of this example, as follows.

[4]: Discovery of a new scientific theory. Now let us relabel D to be a new scientific theory, and $\overline{D}$ to be a familiar old theory. $+$ is the positive outcome of a significance test at significance level 0.99 for rejecting hypothesis $\overline{D}$, where the power of the test under the only alternative, namely D, is also 0.99. The probabilistic specifications are exactly as for case [3]. If you are sceptical of the new theory, a priori, and award it a prior probability of 0.001, then again the Bayes analysis gives posterior probability for D of 0.09. Again, different individuals may react differently, and the sensitivity analysis for the effect of the prior on the posterior is the analysis of the scientific community, so that the answer should now be an interval of posterior values which may be reasonably held by individual scientists. If this interval is wide, then the data has not been sufficient to resolve the scientific issue. However, the posterior interval will usually be smaller than the prior interval. Further, if the prior probability for D is not included, then we cannot express a meaningful posterior judgement about D given the data.

Rating this example on the positive arguments **[C1]**, and **[U1]**, we would argue that the traditional statistical assessment of a positive result on a highly significant, very powerful test is not, of itself, a convincing inference. Conversely, for the individual, the prior assessment of the plausibility of the new hypothesis can be converted into a posterior assessment, and without such an assessment there can be no inference. However, it is depressing how often astonishing scientific advances are announced based on precisely such probabilistic errors, based on the misunderstanding that such conclusions are based on the usual standards of scientific evidence. A conditional probability of data given hypothesis cannot demonstrate anything of itself. In this view, a sensitivity analysis over the reasonable a priori judgements of the scientific community gives the full analysis. Such a sensitivity analysis addresses the issue **[I1]** that the Bayes solution has cheated by introducing arbitrary prior assumptions, and, in part, addresses the question **[H1]** as to exactly how we should formulate our prior judgements. Note further:

(i) Likelihood is fundamental to inference, and we should require our procedures to obey the likelihood principle. However, in many problems the likelihood is as subjective as the prior distribution, and, especially in high dimensions, the likelihood as a point-wise property is often highly non-robust.

(ii) The logic of the Bayes position is that if an individual holds a particular collection of prior beliefs and observes particular data, with a specified likelihood, then that individual should hold the prescribed posterior beliefs. This does not support the existence of "objective" or "non-informative" prior distributions, which in general have no special status and are only useful for illustrative purposes in showing the results of the inference under a particular, usually somewhat interesting, choice of prior, or with large amounts of data which will overwhelm whichever prior distribution is chosen, in which case it is helpful to sidestep the need for detailed prior specification.

3 BAYES ANALYSIS FOR LARGE AND COMPLEX PROBLEMS

How does the Bayesian approach scale up for analysing the types of problems arising in complex physical experiments? A brief description is as follows. We construct a mathematical model for the physical system, often implemented as computer code. The model takes as input certain physical and conceptual parameters, x, some being of direct interest, while others are nuisance parameters. For any values of x, the model produces outputs $s(x)$, corresponding to observable experimental outcomes z. The values of the input parameters are largely unknown. To determine these values, we seek those inputs for which $s(x)$ is in close correspondence to z, subject to (i) random simulation features of the model, (ii) measurement errors in z and (iii) discrepancies between the physical system and computer simulator, arising as the model is a simplification and idealisation of the actual process.

The Bayesian approach is well-suited to address such problems as it can handle in a unified manner all of the different types of uncertainty that arise. In low dimensional versions of the problem, the Bayes solutions are indeed natural and sensible. In more complex formulations, for which the input and output spaces for the model are high dimensional and each evaluation of the model can take a very long time, careful prior specification is very important. However, meaningful prior specification of beliefs in probabilistic form over very large possibility spaces is very difficult and may lead to a lot of arbitrariness in the specification, with corresponding technical difficulties in the subsequent analysis.

Bayes linear methodology is an alternative approach which is similar in spirit to the full Bayesian methodology, but which seeks to simplify the burdens of prior specification and analysis by only requiring prior specification of means, variances and covariances between all of the quantities of interest. An overview of the methodology may be found in [4]. At the simplest level, if we have two random vectors B, D, and we specify prior expectations, variances and covariances for all elements of B, D, then the adjusted mean vector B given observation of D is

$$\mathrm{E}_D(B) = \mathrm{E}(B) + \mathrm{Cov}(B, D)(\mathrm{Var}(D))^{\dagger}(D - \mathrm{E}(D)), \tag{3}$$

where $(\mathrm{Var}(D))^{\dagger}$ is a generalised inverse of $\mathrm{Var}(D)$. Bayes theorem is the special case of the above where the elements of D are the indicator functions for events comprising a partition, and (3) reduces to

$$P(X|D) = \sum_i P(X|D_i)D_i,$$

the random quantity which takes value $P(X|D_i)$, if the outcome D_i is observed. The adjusted variance matrix, for B by D, namely $\mathrm{Var}_D(B) = \mathrm{Var}(B - \mathrm{E}_D(B))$, is given by

$$\mathrm{Var}_D(B) = \mathrm{Var}(B) - \mathrm{Cov}(B, D)(\mathrm{Var}(D))^{\dagger}\mathrm{Cov}(D, B) \tag{4}$$

The foundations for this approach are derived from temporal coherence implications for partially specified collections of prior beliefs, and the approach may be viewed as an appropriate way of handling partially specified uncertainties, with full Bayes analysis as a special case; see [5]. In order to apply a Bayes linear approach to the analysis of physical experiments, we must express our second order (means, variances and covariances) beliefs linking all of the ingredients of the problem. We may do this by specifying second order beliefs over the following three equations. Firstly, the measurement equation links observations z to underlying values y by the addition of independent measurement error ϵ, as

$$z = y + \epsilon \tag{5}$$

Secondly, the discrepancy equation links y to the simulator output s^* for the best system input through addition of independent discrepancy term η (which may be further partitioned into local discrepancies resulting from irregularities of the current experimental setup, and global discrepancies resulting from problems with the theoretical formulation, which therefore correlate apparently unrelated experimental discrepancies at different locations),

$$y = s^* + \eta \tag{6}$$

Finally, beliefs about the simulator output are linked with input through an emulator equation, e.g.

$$s(x) = \beta^T G(x) + \delta(x) \tag{7}$$

where β are unknown constants and $G(.)$ is a collection of known functions, expressing systematic global variation in $s(x)$, and δ is a stationary mean zero process in x, representing local variation.

With the above specification, we may carry out a Bayes linear analysis of the physical model. From the combination of the likelihood, the discrepancy and the emulator, we may construct "plausibility" measures based on evaluations of the current value of $(\mathrm{E}(s(x)) - z)$, standardised by the standard deviation of $(s(x) - z)$. This allows us to screen out nuisance parameters, and identify the plausible ranges for the parameters of interest. This approach can be used to drive an approach to experimental design where sequential choices of experiments and of choices of simulator evaluations is directed by the aim of reducing uncertainty for inputs which are currently plausible. After each evaluation, we carry out diagnostic checks for the emulator and the physical model. If these are acceptable, we update beliefs for the emulator and therefore recompute the plausibility function, from which we choose new model and experimental evaluations. When the plausible region is sufficiently reduced in volume, we refit our emulator to the reduced space and we continue in this way until we find all matches between the parameters and the observations which are consistent with our formulation (or until we run out of time, money or patience). Finally, we make a last diagnostic comparison of observed to expected model discrepancy, which allows us to judge overall match quality and assess the quality of forecasts made by the underlying theory. Much of this methodology is recent; see [2], [3] for development of this approach and [7] for a complementary full Bayes approach.

While the current issues debate that we began by discussing is important, I feel that the ultimate reason why Bayesian methods will achieve widespread acceptance is that they are the only approach that is capable of addressing large and complex problems. In my view, **[C2]** and **[U2]** are borne out through approaches such as the above, as they encapsulate and sensibly apply all of the judgements that are required in order to formulate a meaningful combination of experimental design, data analysis, inference and model diagnostics. It is hard to argue for **[I2]** as there is no obvious alternative within a traditional frequency formulation. Objection **[H2]** is more serious, and the need to tame the complexity of belief specification and analysis is a major research area though, as I have suggested, this complexity may be much reduced, where appropriate, by a Bayes linear approach.

4 CONCLUSIONS

Informally, most scientists think as Bayesians. What is to be gained by formal use of Bayes methodology? Obvious areas are

Experimental design: Enormous sums of money are committed to experiments. Efficient design to optimise information and maximise the chance of valuable discoveries should be driven by decision theoretic formulations built on careful specification of uncertainties.

Analysis: In high dimensional problems, Bayes analysis extracts much more information from the data than traditional approaches, for example by focusing on the key areas of the likelihood surface.

Combining results While the experimenter may be interested in the outcome of a single experiment, the community needs to combine the values arising from a variety of different experiments and theoretical considerations and the Bayesian approach gives a systematic method for such combination.

More generally, the field is ripe for unifying Bayes solutions, based around subjective, rather than objective, Bayesian formulations, using sensitivity analysis to describe the extent to which the scientific community should be brought into consensus by the currently available experimental results.

References

[1] Bernardo, J.M. and Smith, A.F.M. (1994) Bayesian Theory, Wiley, Chichester.

[2] Craig, P. S., Goldstein,M. Seheult, A. H. and Smith, J.A. (1997) Pressure matching for hydocarbon 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. C. Gastonis et al. 37-93. Springer-Verlag.

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

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

[5] Goldstein, M. (1996). Prior inferences for posterior judgements, in Proceedings of the 10th International Congress of Logic, Methodology and Philosophy of Science, M. L. D. Chiara et al eds, Kluwer.

[6] Howson, C. and Urbach, P. (1989) Scientific Reasoning: The Bayesian Approach. Open Court, Chicago.

[7] 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.

[8] O'Hagan, A. (1994) Kendall's advanced theory of statistics 2B: Bayesian inference, London:Edwin Arnold