Bayesian Analysis (2006) **1**, Number 3, pp. 385–402

The Case for Objective Bayesian Analysis

James Berger*

Abstract. Bayesian statistical practice makes extensive use of versions of objective Bayesian analysis. We discuss why this is so, and address some of the criticisms that have been raised concerning objective Bayesian analysis. The dangers of treating the issue too casually are also considered. In particular, we suggest that the statistical community should accept formal objective Bayesian techniques with confidence, but should be more cautious about casual objective Bayesian techniques.

Keywords: History of objective Bayes, reference priors, matching priors, invariance, information, Jeffreys priors, frequentist validation, subjective Bayes, elicitation, unification of statistics, coherency, marginalization paradox, vague proper priors, data dependent priors.

1 Introduction

This is not meant to be a general introduction to objective Bayesian analysis. Indeed, little space is spent in actually defining objective Bayesian analysis, or describing my favorite approaches. The article instead simply addresses the debate as to the value of objective Bayesian versus subjective Bayesian analysis. Note that, in practice, I view both objective Bayesian analysis and subjective Bayesian analysis to be indispensable, and to be complementary parts of the Bayesian vision. But, in this debate, I will be focusing on the strengths of objective Bayes and the weaknesses of subjective Bayes.

Section 1 sets the stage, outlining the framework in which I am considering the debate. Section 2 outlines the case for use of objective Bayesian methods; Section 3 considers the arguments against objective Bayesian analysis; and Section 4 considers the dangers of too-casual objective Bayesian analysis.

1.1 Probability

Probability can be formally derived from a variety of axiom systems, but intuition that people have concerning probability does not arise from any one such system, and so I do not view it to be useful to restrict consideration to a single definition. Thus I will simply view probability as a primitive concept, something that satisfies the standard rules of probability and is used to describe an individual's (or group's) "degree of belief" in the occurrence of an event. This is not as tidy as the subjectivist position on probability, but it allows access to objectivist Bayesianism by many who would not accept the subjectivist definition.

*Duke University and SAMSI, Durham, NC, http://www.stat.duke.edu/~berger

 ba0003

1.2 Philosophical viewpoints

There is no unanimity as to the definition of objective Bayesian analysis, nor even unanimity as to its goal. Indeed, there are at least the following four philosophical positions:

1. A major goal of statistics (indeed science) is to find a completely *coherent* objective Bayesian methodology for learning from data. This is exemplified by the attitudes of Jeffreys (1961) and Jaynes (1999) (although Kass and Wasserman (1996) suggests that the viewpoint of Jeffreys may have gradually changed to reflect something closer to Viewpoint 3 below).

2. Objective Bayesian analysis is the *best* method for objectively synthesizing and communicating the uncertainties that arise in a specific scenario, but is not necessarily coherent in a more general sense.

3. Objective Bayesian analysis is a convention we should adopt in scenarios in which a subjective analysis is not tenable.

4. Objective Bayesian analysis is simply a collection of adhoc but useful methodologies for learning from data.

There is little disagreement as to 4), at least among Bayesians who have done extensive data-analysis. 2) asks for something rigorous – but less so than 1) – namely, definition of optimal objective Bayesian procedures in light of a well-defined scenario. For instance, the scenario might be as specific as "inference for the correlation coefficient from a bivariate normal distribution." One approach in this direction (that I very much like) is the reference prior approach of Bernardo (1979). Another such specific scenario and recommendation is "for a group-invariant problem involving an amenable group, use the right-Haar measure" (see Berger (1985) for a review). 3) suggests the still weaker view that objective Bayesian analysis should just be treated conventionally, with the community formally agreeing to use certain default procedures when subjectivity cannot be incorporated into the prior. For instance, when considering tests or models that have differing dimensions, it seems impossible to rigorously define a *best* method of communication, so that we will likely have to settle for conventional methods.

My general view is that 1) is not attainable; 2) is often attainable and should be done if possible; but often (as in most testing or model selection problems) the best we can hope for is 3).

1.3 Should we be using the word "objective"?

Many Bayesians object to the label "objective Bayes," claiming that it is misleading to say that any statistical analysis can truly be objective. I agree with this at a philosophical level (cf. Berger and Berry (1988)), because the process of data analysis typically involves a host of subjective choices, such as model building, and these choices will typically have a much greater effect on the answer than will such things as choice of prior distributions for model parameters.

Model-building is not typically part of the objective/subjective debate, however – in part because of the historical success of using models, in part because all the major philosophical approaches to statistics use models and, in part, because models are viewed as "testable," and hence subject to objective scrutiny. It is quite debatable whether these arguments are sufficient to remove model choice from the objective/subjective debate, but I will simply follow statistical (and scientific) tradition and do so.

The context of my discussion, therefore, will be the choice of prior distributions for parameters of statistical models. In this domain, I feel that there are a host of practical and sociological reasons to use the label "objective" for priors of model parameters that appropriately reflect a lack of subjective information. Other names for this approach have been suggested – such as noninformative, reference, default, conventional, and non-subjective – but none carries the simplicity or will carry the same weight outside of statistics as "objective." The statistics profession, in general, hurts itself by not using attractive names for its methodologies, and we should start systematically accepting the "objective Bayes" name before it is co-opted by others.

1.4 History

A common misconception is that Bayesian analysis is a subjective theory; this is neither true historically nor in practice. The first Bayesians, Bayes (see Bayes (1763)) and Laplace (see Laplace (1812)), performed Bayesian analysis using a constant prior distribution for unknown parameters, although the motivations of each in doing so were considerably more sophisticated than simply stating that each possibly value of the parameter should receive equal prior weight. Indeed, this approach to statistics, then called "inverse probability" (see Dale (1991)) was central to statistics for most of the nineteenth century, and was highly influential in the early part of the twentieth century. Criticisms of only using a constant prior distribution led to alternative approaches to statistics being developed, such as the frequentist school and the Fisherian school. Objections to these philosophies – primarily based on the notion that statistical actions should be coherent in various senses – resulted in the rapid growth of the subjective Bayesian school of statistics in the last half of the 20th century. Notice, however, that the subjective Bayesian school did not become a serious component of Bayesianism until nearly 200 years of highly successful objective Bayesian practice.

During the rise of the frequentist and Fisherian schools, the objective Bayesian school did not disappear. Indeed, Jeffreys (see Jeffreys (1961)) was introducing significant refinements of the inverse probability theory, refinements which eliminated most of the earlier objections to inverse probability. Most of the applied Bayesian analyses I see today follow the Laplace-Jeffreys objective school of Bayesian analysis, although often with additional modern refinements.

As mentioned above, the most familiar element of the objective Bayesian school is the use of objective prior distributions, designed to be minimally informative in some sense. The most famous of these is the *Jeffreys-rule* prior (see Jeffreys (1961)). *Reference* priors (Bernardo (1979) and Berger and Bernardo (1992)) are a refinement of the Jeffreys-rule priors for higher dimensional problems and have proven to be remarkably successful from both Bayesian and non-Bayesian perspectives. *Maximum entropy* priors (see Jaynes

(1999)) are another well-known type of noninformative prior (although they often also reflect certain informative features of the system being analyzed); related are *minimal description length* and *minimal message length* priors (cf. the Computer Journal: special issue on Kolmogorov complexity and algorithmic information theory, Volume 42, Issue 4: 1999). *Invariance* priors, as mentioned above, *matching* priors (see Datta and Mukerjee (2004)), and *admissible* priors (cf. Berger et al. (2005)) are other approaches being extensively studied today. A number of other approaches, such as *data translated likelihood* priors (cf. Box and Tiao (1973)) and *maximal data information* priors (cf. Zellner (1977)) have been extensively studied in the past. Finally, the *fiducial approach* (cf. Fisher (1930)) and the *structural approach* (cf. Fraser (1968)) were developed as alternatives to objective Bayesian analysis, and typically have objective Bayesian interpretations. See Kass and Wasserman (1996) for a review of methods for selecting noninformative priors.

A quite different area of the objective Bayesian school is that concerned with techniques for default model selection and hypothesis testing. We include some comments about these problems in the article, but they are of a distinct enough nature (and are of enough difficulty technically and conceptually) that we do not focus upon them.

There are two other areas of Bayesian statistics that are, depending on one's point of view, highly related to objective Bayesian analysis. One is *robust Bayesian analysis*, which is discussed in Section 3.3. The other is *nonparametric Bayesian analysis* which, in some sense, attempts to do for model selection what objective Bayesian analysis does for selection of priors for parametric models. Discussion of nonparametric Bayesian analysis would take us too far afield.

2 Motivations for Objective Bayesian Analysis

2.1 The appearance of objectivity is often required

This is at once the heart, and the superficiality, of the objective Bayesian position. Scientists hold up objectivity as the ideal of science, but often fail to achieve it in rather spectacular fashion. This does not give any breathing room for statistics, however, because it is statistics that the scientists often call upon to provide objective validation of what they do. Thus the (arguably correct) view that science should embrace subjective statistics falls on deaf ears; they come to statistics in large part because they wish it to provide objective validation of their science.

Superficiality of the appearance of objectivity is mentioned above because this sociological rationale for objective Bayesian analysis does not reflect the depth of what it actually achieves – a readily understandable communication of the information in the observed data, as communicated through a statistical model, for any scientific question that is posed. In addition, we will argue that it simultaneously achieves what should be a major goal of Bayesianism – ensuring that answers are conditional on the data actually obtained – while at the same time respecting the frequentist notion that the methodology must ensure success in repeated usage by scientists.

Regulatory analysis by government is another regime in which the appearance of objectivity is highly valued. In a different world, clinical trials for a new drug would incorporate the subjective opinions of all scientists involved but, since most of the scientists would be from the interested pharmaceutical company, regulatory agencies instead prefer to base their analyses on objective methodology. There are contexts – such as regulation of medical devices – where it is arguably essential to utilize the subjective Bayesian approach. And there are ethical reasons why it might become necessary for clinical trials to be done in a subjective Bayesian fashion (cf. Kadane (1996)). But keeping the appearance of objectivity, to the extent possible, will likely remain a major goal of regulatory agencies.

Many more scenarios could be listed, but would be superfluous. Statistics owes its central presence in science and life to the facts that (i) it is enormously useful for prediction; (ii) it is viewed as providing an objective validation for science. Discarding the latter would lead to a considerable downgrading of the value placed on statistics.

2.2 Most statistics is not done by statisticians

For whatever reasons, it is not common for subjective Bayesian statistical analyses to be done by non-statisticians. Indeed, one often hears the comment, "I do not want to do a subjective analysis, and hence I will not use Bayesian methodology." This is often a tragedy; one can still take advantage of the many benefits of Bayesian analysis, even if a subjective Bayesian analysis is not performed. Among the many such benefits are:

- Highly complex problems can be handled, via MCMC.
- Very different information sources can easily be combined.
- Multiple comparisons are automatically accommodated.
- Bayesian analysis is an automatic "Ockham's razor," naturally favoring simpler models that explain the data.
- The methodology does not require large sample sizes.
- Sequential analysis (e.g. clinical trials) is much easier.

Here is a typical example of the simplicity of objective Bayesian methods in use by non-statisticians.

Example 1. *Medical diagnosis* (Mossman and Berger (2001)): Within a population for which $p_0 = \Pr(\text{Disease } D)$, a diagnostic test results in either a Positive (+) or Negative (−) reading. Let $p_1 = \Pr(+ \mid \text{patient has } D)$ and $p_2 = \Pr(+ \mid \text{patient does not have } D)$. By Bayes' theorem,

$$\theta = \Pr(D|+) = \frac{p_0\, p_1}{p_0\, p_1 + (1-p_0)p_2}.$$

In practice, the p_i are typically unknown but, for $i = 0, 1, 2$, there are available (independent) data, x_i, having $\text{Binomial}(x_i \mid n_i, p_i)$ densities. It is desired to find a $100(1-\alpha)\%$ confidence set for θ.

This problem has been studied in the classical literature, using standard log odds and delta method procedures to develop confidence sets, as well as more sophisticated approaches such as the Gart-Nam procedure (see Gart and Nam (1988)). For a description of these methods, as applied to this problem, see Mossman and Berger (2001).

In 1999, Dr. Mossman (a psychiatrist), called me to ask if objective Bayesian analysis could provide a simple answer to this question, since the classical approaches were either difficult to apply or were not proving very successful in the contexts in which he was interested. I immediately gave him the following simple objective Bayesian procedure, utilizing the standard Jeffreys-rule priors ($\pi(p_i) \propto p_i^{-1/2}(1-p_i)^{-1/2}$) for the p_i, to compute the $100(1-\alpha)\%$ equal-tailed posterior credible sets for θ.

- Draw random p_i from the $\text{Beta}(p_i \,|\, x_i + \frac{1}{2}, n_i - x_i + \frac{1}{2})$ posterior distributions, $i = 0, 1, 2$, that results from using the Jeffreys-rule priors with the data.
- Compute the associated $\theta = \dfrac{p_0\, p_1}{p_0\, p_1 + (1-p_0)p_2}$.
- Repeat this process $10,000$ times.
- The $\alpha/2$ and $1-\alpha/2$ quantiles of these 10,000 generated θ form the desired confidence limits. (In other words, simply order the 10,000 values of θ, and let the confidence interval be the interval between the $10^4\,(\alpha/2)$-th and $10^4\,(1-\alpha)/2$-th values.)

This is probably not the optimal objective Bayesian analysis (since it did not attempt to find the optimal objective prior, given that θ was the parameter of interest), but it worked very well (as will be seen in the next section).

Dr. Mossman was able to implement this procedure on an Excel spreadsheet with a few hours work, and so he (and psychiatry) now have a simple and easy to use procedure, routinely usable on a host of applications. In contrast, the subjective Bayes approach would require separate analyses for each individual problem, and would require a considerably higher level of training of the psychiatrists (given the difficulties of accurate subjective elicitation). In addition, there would be a considerable barrier to implementation of the subjective Bayesian approach because of the scientific arena in which this is being used.

2.3 Unification of statistics by objective Bayesian methods

Bayarri and Berger (2004) review the vast literature showing that objective Bayesian methods are the most promising route to the unification of Bayesian and frequentist statistics. Consider, for instance, the medical diagnosis example.

Example 1 (continued): Dr. Mossman was not a Bayesian, and indeed wanted a frequentist confidence interval. To convince him that the Bayesian credible interval is actually an excellent frequentist confidence procedure, a large simulation study was performed. Table 1 gives typical results from this simulation for the objective Bayesian method, along with the three frequentist methods mentioned above. It is based on a simulation that repeatedly generates data from binomial distributions with sample sizes 20 and the indicated values of the parameters (p_0, p_1, p_2). For each generated triplet of data, the 95% confidence interval is computed by the studied procedures, and it is noted whether the interval contains the true θ or misses to the left or right. The entries in the table are the long run proportion of misses to the left or right. Ideally, these proportions should be 2.5% and, at the least, their sum should be 5%.

Clearly the objective Bayes interval has better performance in this regard than any of the classically derived confidence intervals. Furthermore, it can be seen that the objective Bayes intervals are, on average, smaller than the classically derived intervals. (See Mossman and Berger (2001) for more extensive computations.) This combination of small confidence sets with correct frequentist coverage has been repeatedly shown to happen when the objective Bayesian approach is used to derive confidence sets.

Table 1: The probability that the nominal 95% interval misses the true θ on the left and on the right, for the indicated parameter values, and when $n_0 = n_1 = n_2 = 20$.

(p_0, p_1, p_2)	O-Bayes	Log Odds	Gart-Nam	Delta
$(\frac{1}{4}, \frac{3}{4}, \frac{1}{4})$	.0286, .0271	.0153, .0155	.0277, .0257	.0268, .0245
$(\frac{1}{10}, \frac{9}{10}, \frac{1}{10})$	.0223, .0247	.0017, .0003	.0158, .0214	.0083, .0041
$(\frac{1}{2}, \frac{9}{10}, \frac{1}{10})$	.0281, .0240	.0004, .0440	.0240, .0212	.0125, .0191

This but scratches the surface of the objective Bayesian and frequentist interface. In Bayarri and Berger (2004), a variety of other aspects of this interface are considered, including a host of other areas in which objective Bayesian methodologies can be used to obtain excellent frequentist procedures, such as hierarchical, multilevel, or mixed model analysis, and problems that involve many nuisance parameters. Also discussed is the role of objective Bayesian analysis in implementation of conditional frequentist analysis, a methodology that can be much superior to traditional unconditional frequentist analysis.

On the other hand, Bayarri and Berger (2004) review some of the many ways that frequentist notions are valuable in the construction of objective Bayesian methodology, including the development of objective prior distributions and the creation of a variety of simplifications based on asymptotic or other approximations.

2.4 Uses of objective Bayesian analysis to a subjectivist

Most subjectivists do make at least some use of objective Bayesian methods in practice. It is worthwhile discussing some of the reasons why this is so.

Use subject matter experts wisely. One only has limited time to elicit models and priors from the experts in a problem, and usually it is most efficient to use the available expert time for modeling, not for prior elicitation. Indeed, in the model construction phase of an analysis, it is usually quite counterproductive to perform subjective prior elicitation about parameters of models, since the models will likely not even be those used at the end.

Another aspect of this is that statistical model building is typically an activity with which experts will have some familiarity, whereas they will typically not have familiarity with prior elicitation. This last means that, with the subjective Bayes approach, the already scarce time must be used to train them in elicitation, for otherwise the priors obtained can be quite bad.

Example 2. My first large-scale Bayesian analysis was in Andrews et al. (1993). It involved a roughly 3000 parameter problem, with many interweaving levels of hierarchical modeling. For a number of these parameters there was effectively no data, so a massive subjective elicitation effort was undertaken. While the elicitation was done with statistically sophisticated engineers, it was enormously difficult and expensive, with extensive training needed. All the usual elicitation mistakes were encountered: variability was initially much too small (virtually never would different experts give prior distributions that even overlapped); there would be massive confusions over statistical definitions (e.g., what does a positive correlation mean?); etc. Since there was no data available about these parameters, one of the severe problems in elicitation was at least avoided, namely how to elicit the prior when the expert has already seen the data (the usual situation that a statistician faces).

For the many parameters for which there was data, however, all of the expert time was used to assist model building. It was necessary to consider many different models, and expert insight was key to obtaining good models; there simply was no extra available expert time for prior elicitation. Also, the analysis was related to government regulation of fuel efficiency, so it was important for the industry to present an analysis that appeared to be as objective as possible (and, in essence, the model parameters for which objective priors were used were the most contentious of the parameters politically).

This analysis employed one of the largest subjective elicitations I have seen, yet it was still the case that the vast majority of parameters in the problem were handled with objective priors. In subsequent big problems, I have tended to be forced, just from these practical considerations, to use an even higher ratio of objective to subjective priors.

Determining if subjective elicitation is needed. An objective Bayesian analysis can be run initially, to assess if subjective priors are even needed. (Perhaps the data will "swamp the prior.")

Objective Bayes as a reference. Subjective elicitation can easily result in poor prior distributions, because of systematic elicitation bias (*e.g.*, the almost universal tendency to underestimate the actual amount of uncertainty about unknowns) and the fact that elicitation typically yields only a few features of the prior, with the rest of the prior (*e.g.*, its functional form) being chosen in a convenient, but possibly inappropriate, way. It is thus good practice to compare answers from a subjective analysis with answers from an objective prior analysis. If there are substantial differences, it is important to check that the differences are due to features of the prior that are trusted, and not due to either unelicited "convenience" features of the prior or suspect elicitations (*e.g.*, too-small prior variances).

Use for nuisance parameters. A common and reasonable practice is to develop subjective priors for the important parameters or quantities of interest in a problem, with the unimportant or "nuisance" parameters being given objective priors.

Inappropriateness of standard (e.g., conjugate) priors. Through study of objective priors, one can obtain insight into possibly bad behavior of standard (*e.g.*, conjugate) subjective priors. An example is use of the Inverse Wishart distribution as a conjugate prior for a covariance matrix. Yang and Berger (1994) show a very unsettling property of this prior, namely that it forces apart eigenvalues of the covariance matrix.

2.5 Teaching of elementary statistics is greatly simplified

Most elementary statistical procedures have an objective Bayesian interpretation (and indeed many were first derived in the inverse probability days of objective Bayesianism). Teaching the procedures with this interpretation is much easier than teaching them with frequentist interpretations: it is quite a bit easier to understand "θ is in the interval $(2.31, 4.42)$ with degree-of-belief probability 0.95" than to understand "the confidence procedure $C(x)$ will contain θ with probability 0.95 if it were repeatedly used with random data from the model for a fixed θ, and the interval for the given data happened to be $(2.31, 4.42)$." Teaching objective Bayesian analysis is also considerably easier than teaching subjective Bayesian analysis, in that one does not need to teach the difficult subject of elicitation, which requires significant understanding of probability. Note that I am referring here to courses where students are not expected to be able to derive procedures; where the goal is simply to teach then to use the procedures and understand the interpretation of the answers.

3 Arguments Against Objective Bayesian Analysis

3.1 Impropriety

Objective Bayesian procedures often utilize improper prior distributions, which can lead to the worry that one is using "improper probability theory." The major objective Bayesian theories address this, however, by establishing that the posterior distributions that result are proper, and can be justified as limiting approximations to proper prior posteriors.

The reference prior approach seems particularly attractive from this regard, virtually always yielding probabilistically justifiable proper posterior distributions. The Jeffreys-rule prior is usually also successful in this regard. In contrast, the constant prior will fairly often result in improper posteriors. Why the reference and Jeffreys-rule approaches succeed in this regard is not well understood.

3.2 Model and criterion dependence

The major objective Bayesian schools require consideration of the statistical model in choice of the objective prior. For instance, the reference prior school defines the prior via the (asymptotic) model-averaged information difference between the prior and the posterior; the matching prior approach seeks priors that yield optimal frequentist confidence sets for the given model, and the invariance approach is a model-dependent concept. Having priors depend on the model can lead to violations of basic principles, such as the likelihood principle and the stopping rule principle (cf. Berger and Wolpert (1984)).

The most common response of objective Bayesians is that "objectivity" can only be defined relative to some frame of reference, and the natural frame for statistics is the statistical model. Hence violation of principles such as the likelihood principle is the price that has to be paid for objectivity. In practice, violations of the likelihood principle also tend to be extremely minor, with the answers (for the same likelihood function) being virtually the same from one model to another.

It is also the case that objective priors can vary depending on the goal of the analysis for a given model. For instance, in a normal model, the reference prior will be different if inference is desired for the mean μ or if inference is desired for μ/σ. This, of course, does not happen with subjective Bayesianism. Again, the objective Bayesian responds that objectivity can only be defined relative to a frame of reference, and this frame needs to include the goal of the analysis.

3.3 Incoherency

Because objective Bayesian methods can violate principles such as the likelihood principle, they can be incoherent according to standard definitions of coherence. It is worthwhile to discuss what this does, and does not, mean.

There have been a huge variety of axiomatic systems that seek to define coherent inference or coherent decision making, and they all seem to lead to some form of Bayesianism. The typical conclusion of these systems is that the analysis must be compatible with some form of subjective Bayesianism to be fully coherent. At the end of this section, however, we examine whether the usual forms of subjective Bayesianism are really coherent in practice.

Robust Bayesian analysis is a coherent system that is implementable in practice. In this approach subjective specifications are intervals (e.g. $P(A) \in (0.45, 0.5)$), and the analysis considers all priors compatible with these specifications and results in intervals as the posterior conclusions. The foundational arguments for robust Bayesian analysis are compelling (cf. Walley (1991)) and there is an extensive literature on the development of robust Bayesian methodology, much of it reviewed in Berger (1994) and Ríos Insua and Ruggeri (2000). Unfortunately, robust Bayesian analysis has not caught on in statistics, primarily because of technical issues: the interval specifications that are easy to work with seem to result in posterior intervals that are too wide for use in practice, and refining the class of prior distributions to eliminate unreasonable priors leads to computational challenges that limit applicability. This could change as understanding and computation advances, but robust Bayesian analysis is currently much less utilized than objective Bayesian analysis.

A quite different coherent subjective Bayesian approach is the *Bayes linear analysis* approach (cf. Goldstein (1999)) which is based on making expectation, rather than probability, the primitive in dealing with uncertainty. Modulo this assumption (which practically implies that one approaches subjective Bayesian analysis through elicitation of means, variances and covariances of direct quantities of interest, as opposed to the standard statistical use of models and priors on the parameters of models), the approach has appeal in many contexts. Discussion of the strengths and weaknesses of this approach would be an article in itself, so we will remain focused on the more traditional subjective Bayesian analyses here. Indeed, any subsequent reference to "subjective Bayesian analysis" should be understood as not necessarily referring either to robust Bayesian analysis or Bayes linear analysis.

Another issue is that of properly defining coherency when the goal of the analysis is communication of information. Let's consider this question in terms of two commonly mentioned "coherency violations" of objective Bayesian analysis.

Betting incoherency is one of the very common forms of coherency; the idea is to ensure that any stated probability specifications cannot be beaten by a "Dutch book," i.e., an opponent cannot formulate a betting strategy according to which s/he will for sure win money from you (in some repetitive version of the game) if you were to bet according to your probability specifications. This seems reasonable, but consider the simple example of observing data $x \sim N(\theta, 1)$, where the goal is to objectively communicate the probability that $\theta < x$. Almost any objective approach (Bayesian, frequentist, fiducial) would say the objective probability is 1/2, but this is betting incoherent (see Robinson (1979)). Betting incoherency thus seems to be too strong a condition to apply to communication of information.

Marginalization paradoxes can arise from objective priors, as shown in Dawid et al. (1973). Here is an example, from Berger and Sun (2006), which raises the interesting issue of whether avoidance of a marginalization paradox is fundamentally necessary.

Example 3. The bivariate normal distribution of (x_1, x_2) has mean (μ_1, μ_2) and covariance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where ρ is the correlation between x_1 and x_2. For a sample $\mathbf{x} = \{(x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1n}, x_{2n})\}$, the sufficient statistics are $\overline{x} = (\overline{x}_1, \overline{x}_2)'$, where $\overline{x}_i = n^{-1} \sum_{j=1}^n x_{ij}$, and

$$\mathbf{S} = \begin{pmatrix} s_{11} & r\sqrt{s_{11}s_{22}} \\ r\sqrt{s_{11}s_{22}} & s_{22} \end{pmatrix},$$

where $s_{ij} = \sum_{k=1}^n (x_{ik} - \overline{x}_i)(x_{jk} - \overline{x}_j), \quad r = s_{12}/\sqrt{s_{11}s_{22}}$.

An interesting objective prior here is the *right-Haar* prior (corresponding to the triangular group) $\pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \sigma_1^{-2}(1-\rho^2)^{-1}$. The resulting posterior distribution for ρ, $\pi(\rho \mid \mathbf{x})$, can be written *constructively* (i.e., in a way that simulation from the posterior is straightforward) as

$$\psi\left(-\frac{Z^*}{\sqrt{\chi^{2*}_{n-1}}}+\frac{\sqrt{\chi^{2*}_{n-2}}}{\sqrt{\chi^{2*}_{n-1}}}\frac{r}{\sqrt{1-r^2}}\right), \quad \text{where} \quad \psi(x)=\frac{x}{\sqrt{1+x^2}}, \tag{1}$$

Z^* is a standard normal random variable, and χ^{2*}_{n-1} and χ^{2*}_{n-2} are chi-squared random variables with the indicated degrees of freedom, all random variables being drawn independently to obtain a sample from the posterior.

This situation yields a marginalization paradox, as follows:

- Proper priors have the following intuitively appealing property in multiparameter problems: if the posterior for a single parameter ρ depends only on a statistic r, and the density of r depends only on ρ, then the posterior density of ρ is proportional to the density of $r \mid \rho$ times the marginal prior for ρ. If an improper objective prior violates this property, it is stated to yield a marginalization paradox.
- From (1), it is clear that the posterior $\pi(\rho \mid \mathbf{x})$ depends only on r. ($\pi(\rho \mid \mathbf{x})$ also happens to be the fiducial density for ρ, as found by Fisher (1930), although this does not relate to the issue of marginalization paradoxes.)
- It is a fact that the density of r, $p(r \mid \rho)$, depends only on the parameter ρ.
- It was shown in Brillinger (1962) that there exists no prior density $\pi(\rho)$, such that $\pi(\rho)\, p(r \mid \rho) \propto \pi(\rho \mid \mathbf{x})$, so that the objective Bayes posterior/fiducial density for ρ cannot be produced in a Bayesian way starting just from r, resulting in a marginalization paradox.

This example of a marginalization paradox is interesting because, in all other respects, the posterior $\pi(\rho \mid \mathbf{x})$ seems exemplary. It arises from a highly recommended prior distribution (the right-Haar prior), yields the fiducial distribution for ρ, and also yields exact frequentist inferences. For instance, Bayesian $100(1-\alpha)\%$ credible sets, $C(r)$, for ρ, when considered as frequentist confidence intervals, can be shown to have exact frequentist coverage of $1-\alpha$. It is, at the very least, surprising that an analysis that is simultaneously correct from Bayesian, frequentist, and fiducial perspectives is incoherent according to the marginalization paradox.

Note that some approaches to objective Bayesian analysis, such as the reference prior approach, do seem to avoid the marginalization paradox. For instance, Bayarri (1981) shows that the reference prior for ρ in the bivariate normal example is $\pi(\mu_1,\mu_2,\sigma_1,\sigma_2,\rho) = \sigma_1^{-1}\sigma_2^{-1}(1-\rho^2)^{-1}$, and this does not have a marginalization paradox. But this prior does not result in posterior credible intervals that have exact frequentist coverage, so there is a fascinating philosophical tension.

Is subjective Bayesian analysis coherent in practice? The idealization of subjective Bayesian analysis is of course coherent, but this idealization is not implementable, except for trivial versions such as *always* estimate $\theta \in (0,\infty)$ by 17.35426 (a coherent rule, but not one particularly attractive in practice). The problem is that, to elicit all features of a subjective prior $\pi(\theta)$, one must infinitely accurately specify a (typically) infinite number of things. In practice, only a modest number of (never fully accurate) subjective elicitations are possible, so practical Bayesian analysis must somehow construct the entire prior distribution $\pi(\theta)$ from these elicitations. Is there a coherent way of turning these elicitations into full prior distributions? Consider some standard possibilities.

Example 4. One common method of turning elicitations into full priors is to use conjugate priors. But conjugate priors are model-dependent (or at least likelihood-dependent) so, depending on the experiment designed to study θ, the subjective Bayesian following this "prior completion" strategy would be constructing different priors for the same θ, clearly incoherent.

Example 5. Consider the scientist who is taught to elicit prior quantiles for θ and to fit these to some prior distribution $\pi(\theta)$. Alternatively, the scientist might have been told to elicit predictive distributions, and work back to the prior. Different answers would virtually certainly result from application of these two techniques.

Example 3 (continued). Consider again the above marginalization paradox involving ρ, and suppose the analyst has access to the subject-matter expert for subjective prior elicitation for 3 hours. In case one, the full bivariate normal distribution is considered, and the subjective elicitation is performed for all five parameters. Alternatively, suppose the analyst used only the model $p(r \mid \rho)$, so all the time was used to elicit the prior for just ρ. Almost certainly the subjective Bayesian analyses would have given different answers. So, in practice, subjective Bayesians will virtually always experience what could be called practical marginalization paradoxes.

3.4 The multiplicity of objective Bayesian procedures

In Section 1.4, a few of the many methods for developing objective priors were mentioned. Inventing a new criterion for finding "the optimal objective prior" has proven to be a popular research pastime, and the result is that many competing priors are now available for many situations. This multiplicity can be bewildering to the casual user.

I have found the reference prior approach to be the most successful approach, sometimes complemented by invariance considerations as well as study of frequentist properties of resulting procedures. Through such considerations, a particular prior usually emerges as the clear winner in many scenarios, and can be put forth as the recommended objective prior for the situation. Berger et al. (2006), which is under preparation, will be presenting such unique recommended objective priors.

4 Dangers of Casual Objective Bayesian Analysis

One of the mysteries of modern Bayesianism is the lip service that is often paid to subjective Bayesian analysis as opposed to objective Bayesian analysis, but then the practical analysis actually uses a very adhoc version of objective Bayes, including use of constant priors, vague proper priors, choosing priors to "span" the range of the likelihood, and choosing priors with tuning parameters that are adjusted until the answer "looks nice." I call such analyses *pseudo-Bayes* because, while they utilize Bayesian machinery, they do not carry with them any of the guarantees of good performance that come with either true subjective analysis (with a very extensive elicitation effort) or (well-studied) objective Bayesian analysis. I will briefly discuss the problem with each of these pseudo-Bayes procedures.

It should first be noted, however, that the pseudo-Bayes approach can be extremely effective in actual data analysis, and I do not mean to discourage this approach. It simply must be realized that pseudo-Bayes techniques do not carry the guarantees of proper subjective or objective Bayesian analysis, and hence must be validated by some other route.

4.1 Use of the constant prior density

The initial criticisms of use of a constant prior density in inverse probability centered on the inconsistency of using the same prior for any parameterization of the model. This logical issue is rarely a serious practical concern, but there can be serious practical difficulties with use of a constant prior density.

One such problem is that constant prior densities can often result in improper posterior distributions, as was shown for a common spatial situation in Berger et al. (2001). Another example is discussed in Berger et al. (2005), in which use of a constant density for a covariance matrix requires twice as many normal observations for posterior propriety as the problem should require. This issue is worsened by the common use today of Markov chain Monte Carlo (MCMC) methods of computation, since they can make it difficult to identify improper posterior distributions. In contrast, recall that reference priors virtually always result in proper posterior distributions.

4.2 Vague proper priors

Vague conjugate priors. In general, when an improper prior produces an improper posterior, using a vague proper prior can only hide—not solve—the problem. For instance, in normal hierarchical models with a "higher level" variance τ^2, it is quite common to use the vague proper prior density $\pi(\tau^2) \propto \tau^{-2(\varepsilon+1)} \exp(-\varepsilon'/\tau^2)$, with ε and ε' small. However, as $\varepsilon \to 0$ it is typically the case in these models that the posterior distribution for τ^2 will pile up its mass near 0, so that the answer can be ridiculous if ε is too small. An objective Bayesian, who incorrectly used the related prior $\pi(\tau^2) \propto \tau^{-2}$, would typically become aware of the problem, since the posterior would not converge (as it will with the vague proper prior). The common perception that using a vague proper prior is safer than using improper priors, or conveys some type of guarantee of good performance, is simply wrong.

Use of vague proper priors will work well only when the vague proper prior is a good approximation to a good objective prior. In the hierarchical situation mentioned above, for instance, it is known that $\pi(\tau^2) \propto \tau^{-1}$ is a good objective prior (cf. Berger and Strawderman (1996)), and a good proper prior approximation to this objective prior is $\pi(\tau^2) \propto \tau^{-(\varepsilon+1)} \exp(-\varepsilon'/\tau^2)$, which is an inverse gamma distribution.

Truncation of the parameter space. It is a related common misconception that, to avoid potential difficulties with improper priors, one need only choose (extreme) bounds on the parameter space and confine analysis to this bounded space (in which the prior will presumably be proper). For instance, a common attempt to avoid possible posterior impropriety when using the constant prior is to choose the prior to be constant over some (large) bounded region of Θ. This will not solve the problem, however, in that if the posterior resulting from the constant prior were improper, then the ensuing inferences will often be highly dependent on the actual bounds that were used. (The answers obtained by truncating at $\pm K$ could then be very different than the answers obtained by truncating at $\pm 2K$.) At the very least, this approach should only be used if a very careful sensitivity study is done with respect to these bounds (and with bounds for different parameters varying independently in the sensitivity study).

4.3 Transformation of θ

This approach consists of first transforming θ to a bounded interval – for instance, by defining $\psi = g(\theta)$, where $g : \Theta \to [0, 1]$ is a $1-1$ differentiable transformation – and then placing an objective (but proper) prior on ψ, say $\pi(\psi) = 1$. This will typically ensure that the posterior is proper. Note, however, that choosing g is essentially equivalent to choosing the original prior $\pi(\theta)$. For instance, a simple change of variables shows that (under mild conditions) use of $\pi(\psi) = 1$ is equivalent to use of $\pi(\theta) = |g'(\theta)|$, where $g'(\theta)$ is the derivative of $g(\theta)$. This approach is thus essentially equivalent to choosing a particular proper prior, but is guilty of hiding the fact that this has been done!

4.4 Data-Dependent Priors

We have seen that objective Bayesian priors usually depend on the model chosen for the analysis, but it is not uncommon to see priors that actually depend on the data. These include empirical Bayes priors, vague-proper priors specifically chosen to "span the range of the likelihood function," and objective priors chosen conditional on a local ancillary (*e.g.*, Fraser and Yuan (2004)). This last is an interesting new developing theory of objective priors and, while data dependent, it does not involve an inappropriate double use of the data, which is the problem that arises in the other two approaches.

Empirical Bayes priors: The empirical Bayes approach consists of somehow modeling the prior, with the prior model having unknown *hyperparameters* that are estimated by data (usually by maximum likelihood) and used in the ensuing Bayesian analysis. Philosophically, this results in an undesirable double use of the data, but it is often not a severe double use of the data unless the sample sizes are fairly small. More problematic is that estimates of the hyperparameters can easily be on the boundaries of their parameter space. In variance components problems for instance, it is quite common that the empirical Bayes estimate of a variance component is zero, which can severely and inappropriately affect the ensuing analysis. Thus we strongly prefer full objective hierarchical Bayesian analysis to empirical Bayesian analysis, unless it has been shown that the empirical Bayes analysis does approximate the full Bayesian analysis (see, e.g., Kass and Steffey (1989)).

Data-dependent vague proper priors. The second common data-dependent procedure is to choose priors that span the range of the likelihood function. For instance, one might choose a uniform prior over a range that includes most of the mass of the likelihood function, but that does not extend too far (thus hopefully avoiding the problem of using a "too vague" proper prior). Another version of this procedure is to use conjugate priors, with parameters chosen so that the prior is spread out somewhat more than the likelihood function, but is roughly centered in the same region. The two obvious concerns with these strategies are that (i) the answer can still be quite sensitive to the spread of the rather arbitrarily chosen prior; and (ii) centering the prior on the likelihood is a quite problematic double use of the data. Also, in problems with complicated likelihoods, it can be very difficult to implement this strategy successfully.

Perhaps the most questionable of all the pseudo-Bayes procedures is to write down proper (often conjugate) priors with unspecified parameters, and then to treat these parameters as "tuning" parameters to be adjusted until the answer "looks nice." At the very least, anyone using this technique should clearly explain that this is what was done.

In conclusion, while these pseudo-Bayesian techniques can be useful as data exploration tools, they should not be confused with formal objective Bayesian analysis, which has very considerable extrinsic justification as a method of analysis.

Acknowledgments

This work was supported in part by the National Science Foundation Grant DMS-0103265.

References

Andrews, R. W., Berger, J. O., and Smith, M. H. (1993). "Bayesian estimation of manufacturing effects in a fuel economy model." *Journal of Applied Econometrics*, 8: 5–18. 392

Bayarri, M. J. (1981). "Inferencia Bayesiana sobre el coeficiente de correlacin de una poblacin normal bivariante." *Trabajos de Estadistica e Investigacion Operativa*, 32: 18–31. 396

Bayarri, M. J. and Berger, J. (2004). "The interplay between Bayesian and frequentist analysis." *Statistical Science*, 19: 58–80. 390, 391

Bayes, T. (1763). "An essay towards solving a problem in the doctrine of chances." *Philosophical Transactions of the Royal Society of London, 53, 370-416; 54, 296-325*. 387

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis (Second edition)*. Springer-Verlag: New York. 386

— (1994). "An overview of robust Bayesian analysis, with Discussion." *Test*, 3: 5–124. 394

Berger, J. O. and Bernardo, J. M. (1992). "On the development of reference priors (with Discussion)." In *Bayesian Statistics 4*, 35–60. Oxford Univ. Press: London. 387

Berger, J. O., Bernardo, J. M., and Sun, D. (2006). *Objective Bayesian Inference*. Under preparation.

Berger, J. O. and Berry, D. A. (1988). "The relevance of stopping rules in statistical inference (with Discussion)." In *Statistical Decision Theory and Related Topics IV, Volume 2*, 29–72. Springer-Verlag: New York. 386

Berger, J. O., De Oliveira, V., and Sanso, B. (2001). "Objective Bayesian analysis of spatially correlated data." *Journal of the American Statistical Association*, 96: 1361–1374.

Berger, J. O., Strawderman, W., and Tang, D. (2005). "Posterior propriety and admissibility of hyperpriors in normal hierarchical models." *Annals of Statistics*, 33: 606–646. 388

Berger, J. O. and Strawderman, W. E. (1996). "Choice of hierarchical priors: Admissibility in estimation of normal means." *Annals of Statistics*, 24: 931–951.

Berger, J. O. and Sun, D. (2006). "Objective priors for a bivariate normal model." Technical report, ISDS, Duke University. 395

Berger, J. O. and Wolpert, R. L. (1984). *The Likelihood Principle*. Institute of Mathematical Statistics: Hayward. 394

Bernardo, J. M. (1979). "Reference posterior distributions for Bayesian inference (with Discussion)." *Journal of the Royal Statistical Society, Series B*, 41: 113–147. 386, 387

Box, G. and Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley: Reading. 388

Brillinger, D. R. (1962). "Examples bearing on the definition of fiducial probability with a bibliography." *Annals of Mathematical Statistics*, 33: 1349–1355. 396

Dale, A. (1991). *A History of Inverse Probability*. Springer-Verlag: New York. 387

Datta, G. and Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics*. Springer: New York. 388

Dawid, A. P., Stone, M., and Zidek, J. V. (1973). "Marginalization paradoxes in Bayesian and structural inference (with discussion)." *Journal of the Royal Statistical Society, Series B*, 35: 189–233. 395

Fisher, R. A. (1930). "Inverse probability." *Proceedings of the Cambridge Philosophical Society*, 26: 528–535. 388, 396

Fraser, D. (1968). *The Structure of inference*. Wiley: New York. 388

Fraser, D. and Yuan, X. (2004). "Neutral priors." Technical report, Statistics Department, University of Toronto.

Gart, J. and Nam, J. (1988). "Approximate interval estimation of the ratio of binomial parameters: A review and corrections for skewness." *Biometrics*, 44: 763–770. 390

Goldstein, M. (1999). "Bayes linear analysis." In *Encyclopedia of Statistical Sciences, update volume 3*, 29–34. S. Kotz, et. al. (Eds.). Wiley: New York. 395

Jaynes, E. (1999). *Probability Theory: The Logic of Science*. accessible at the website http://bayes.wustl.edu/etj/prob.html. 386, 387

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press, London. 386, 387

Kadane, J. e. (1996). *Bayesian Methods and Ethics in a Clinical Trial Design*. Wiley: New York. 389

Kass, R. E. and Steffey, D. (1989). "Approximate Bayesian inference in conditionally independent hierarchical models." *Journal of the American Statistical Association*, 84: 717–726.

Kass, R. E. and Wasserman, L. (1996). "The selection of prior distributions by formal rules (Corr: 1998V93 p412)." *Journal of the American Statistical Association*, 91: 1343–1370. 386, 388

Laplace, P. (1812). *Théorie Analytique des Probabilités*. Courcier. 387

Mossman, D. and Berger, J. O. (2001). "Intervals for post-test probabilities: a comparison of five methods." *Medical Decision Making*, 21: 498–507. 389, 390, 391

Ríos Insua, D. and Ruggeri, F. (2000). *Robust Bayesian Analysis*. Springer-Verlag: New York. 394

Robinson, G. K. (1979). "Conditional properties of statistical procedures." *Annals of Statistics*, 7: 742–755. 395

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall: New York. 394

Yang, R. and Berger, J. O. (1994). "Estimation of a covariance matrix using the reference prior." *Annals of Statistics*, 22: 1195–1211. 393

Zellner, A. (1977). "Maximal data information prior distributions." In *New Developments in the Applications of Bayesian Methods*, 211–232. A. Aykac and C. Brumat, eds. North-Holland: Amsterdam. 388