DISCUSSION

J. K. Ghosh and Tapas Samanta

Indian Statistical Institute

A number of objective Bayesian methods of model selection and testing a sharp null hypothesis have been proposed and developed to deal with the difficulties with improper noninformative priors. Professors Berger and Pericchi have been the leading contributors to this area. Their present work is an illuminating overview of the subject as it stands now. As part of the overview they take up the important problem of evaluating these methods and give some final recommendations. We generally agree with them but feel many points need further study.

1. *Intrinsic Priors.* Berger and Pericchi are guided by the principle that a good method should produce a Bayes factor that is equal up to $o_p(1)$ to a Bayes factor with "reasonable default prior". The default Bayes factors like the intrinsic Bayes factor (IBF) and the fractional Bayes factor (FBF) seem to have this correspondence (at least asymptotically) as illustrated with a number of examples.

A potential problem is that it may not be easy to agree to a "reasonable default prior" in examples which have not been enriched by contextual discussions like that of Jeffreys. In such cases we would suggest that one can go a step further and argue as follows. If there is a prior which gives rise to a Bayes factor that is well approximated by an appealing data analytic procedure, then each of the two – the prior and the method – lends support to the other. We feel the above default Bayes factors are naturally developed, intuitively very appealing and may be considered to be good default Bayes factors in their own merits. The intrinsic priors may therefore be considered to be natural default priors as they correspond to naturally developed good default Bayes factors. In turn this argument strengthens the default Bayes factors as argued by Berger and Pericchi.

Thus these automatic methods represent one way of generating (conventional) default priors for hypothesis testing and model selection problems at least in the cases where the model dimension is small compared with sample size.

It is interesting to observe that the Cauchy prior recommended by Jeffreys for the problem considered in Illustration 1 can be obtained in this way, vide Ghosh and Samanta (2001, Section 2.5).

'J. K. Ghosh is Jawharlal Nehru Professor and Tapas Samanta is Professor, Indian Statistical Institute, 203 B.T. Road, Calcutta 700 037, India; email: tapas@isical.ac.in. J. K. Ghosh is also Professor, Department of Statistics, Purdue University, U.S.A; email: ghosh@stat.purdue.edu

The case of large model dimension may need further study, since the approximation by BIC breaks down, vide Berger et al. (1999).

A related technical question is whether an intrinsic prior is a probability measure. The authors have an interesting earlier result on this (Berger and Pericchi 1996a, Theorem 1) which is, however, based on the assumption that $\pi_1^N(\boldsymbol{\theta}_1)$ is proper. If one proceeds in the same way for two parameters or parameter vectors (ξ, η) where η is the parameter of interest, the conditional distribution $\pi_2^I(\eta|\xi)$ will be proper if

$$\int \pi_2^I(\xi, \eta) d\eta < \infty.$$

Using the argument in the proof of Theorem 1 of Berger and Pericchi (1996a), this condition is equivalent to the integrability of

$$\frac{m_1^N(x(l))}{m_2^N(x(l))} m_2^N(x(l), \xi)$$

with respect to $dx(l)$, where

$$m_2^N(x(l), \xi) = \int \pi_2^N(\xi, \eta) f_2(x(l)|\xi, \eta) d\eta.$$

This condition cannot be simplified any further. So a more general result than Theorem 1 of Berger and Pericchi (1996a) would not be easy to obtain. Hence it would be good to have a catalogue of examples where the intrinsic prior is a probability measure or close to being one in some sense. This has been verified by the authors and their students for a number of examples including normal linear models and some exponential models but other examples are worth studying.

2. *Effect of Using Minimal Training Samples.* IBF uses minimal sample for training, leaving as much of the data as possible for model comparison. Use of larger training samples is expected to correspond to Bayes factors with more peaked (intrinsic) priors. This can be easily seen in simple examples like those in Illustrations 1 and 3. In Illustration 3, for example, the intrinsic prior based on minimal training samples (of size 1) is a $N(0, 2)$ prior, whereas training samples of size 2 will lead to a $N(0, 1)$ prior. An argument as to why it is expected to be so in the general nested case is given in Ghosh and Samanta (2001, Section 3).

3. *Scale of Priors.* To fix ideas consider only a one-dimensional parameter θ. Most Bayesian tests and model selection procedures use a prior under the more complex model which has a scale comparable with the scale of variation of a single random variable X_i, namely, σ. This idea goes back to Jeffreys and is usually right. But one can think of special situations where other scales may be more appropriate. For example, if the sample size is chosen carefully at the planning stage, the alternatives like $\theta = \delta/\sqrt{n}$ may

be of special importance. A prior that does not recognize this will put too small a weight on such θ's. This point has been argued in more detail in Ghosh and Samanta (2001, Section 5.2) where it has also been suggested that the apparently dramatic difference between common classical and common Bayesian tests is really a matter of judgement on the scale in a particular problem and that there is no single scale that is right in all problems.

For simplicity, let θ be a location parameter and suppose there is some reason to believe that the prior under M_2 should have a scale comparable with $\sigma/\sqrt{c}$. In this case one may first find an intrinsic prior π with training sample size equal to one and then choose π_c as the distribution of $\theta = \theta'/\sqrt{c}$ where $\theta' \sim \pi$. The prior π_c will be the appropriate prior to use for θ.

If the intrinsic prior π is not easy to find, how can one directly modify the algorithm of IBF's to utilize the information that $\sigma/\sqrt{c}$ is the right scale for θ under M_2? Changing the parameter but keeping the training samples independent of c doesn't seem to help. At least for c growing at a sufficiently slow rate with the sample size n the following recipe should work. The parameter isn't changed but the training sample is of size $[c]$, the integral part of c. For $N(\mu, 1)$ it can be verified this works for fixed c as well as for $c = o(n)$. If $c = O(n)$, one can easily see that this method does not work.

Another different way of handling this problem is to start with a data dependent prior (Ghosh and Samanta 2001, Section 2.5) and make a suitable change there. We have not explored the consequences of doing this.

4. *Examples.* Since this is an exciting area of new research, there is a need to try out these tools in many examples and examine carefully the inference resulting from these tools. There are several examples of this sort in the paper. They are all quite interesting. However, the conclusions regarding the IBF's and FBF that are drawn from them need further study and verification. Justifying these new Bayes factors is difficult enough. To compare them or choose one from among them is even more difficult. It is true that the IBF's are closer cousins of the non-controversial conditional BF's than the FBF and so seem easier to interpret and accept. However, some of the examples provide at best ambiguous evidence about which of these BF's one should use. Some of our reservations are discussed below.

To start with, it seems unclear whether one should compare a single criterion like the FBF with all the IBF's together.

Secondly, in some examples there are aspects that are missed. For example, the fact that the GIBF is smaller than the FBF, vide Section 4.6, doesn't seem to be a strong argument against the FBF. There are several ways of seeing this. In Ghosh and Samanta (2001) we derive the FBF as a GIBF adjusted for the fact that the GIBF uses a data dependent prior that integrates to less than one. Also, if one criterion is always smaller than the other, then it is clear that none of them can be better than the other in all circumstances. It is easy to provide numerical evidence of this sort in Example 6.

Finally, in Example 6, the advantage that the AIBF has over the FBF is bought at a price. For, being exactly equal to a BF with respect to a prior may be preferable to being so only asymptotically.

We too have our own preferences, namely, the AIBF, a trimmed AIBF and the median IBF but believe it's too early to come to any definite conclusion.

Example 1 (Section 4.1) illustrates the difficulties with the class of models with "improper likelihoods" such as the mixture models for which the IBF and FBF cannot be directly employed. It refers to an unpublished work of Shui (1996) that considers modifications of the IBF and the FBF approaches to deal with the mixture models. Hopefully, the interesting work of Shui would be published soon.

5. *Teaching Nonsubjective Bayes Testing.* How should one teach nonsubjective Bayes testing in an undergraduate course? What would be the best way to communicate these ideas to the students who cannot be expected to understand all the subtleties of an IBF? At least in the classical examples ($N(\mu, 1)$ or $N(\mu, \sigma^2)$) it may be easier to motivate and use a BF based on a default or intrinsic prior but one would still have to motivate the prior. Do Berger and Pericchi have any suggestions?

Fulvio De Santis
Università di Roma, "La Sapienza"

1. Introduction.

Model selection and hypothesis testing are difficult topics. In these problems, as we depart from the usual assumption of explicitly stated underlying model, fundamental statistical principles (e.g., likelihood, sufficiency, etc.) begin to fade and we are left with no clear direction. Substantial debate over the appropriate model selection and hypothesis testing problems has taken place inside and outside the Bayesian community. Within the Bayesian approach, controversies arise on how model selection should be performed, even in the *ideal* situation where prior information is available. For example, the recent renewed interest in the development of *default* model selection methods has

[0]Fulvio De Santis is Assistant Professor, Dipartimento di Statistica, Probabilità e Statistiche Applicate , Università di Roma, "La Sapienza", P.le A. Moro, 5, 00185 - Roma, Italy; email: fulvio.desantis@uniroma1.it