Sankhyā: The Indian Journal of Statistics
Special Issue on Bayesian Analysis
1998, Volume 60, Series B, Pt. 1, pp. 1-18

ACCURATE AND STABLE BAYESIAN MODEL SELECTION: THE MEDIAN INTRINSIC BAYES FACTOR*

By J.O. BERGER
Duke University, Durham, USA
and
L.R. PERICCHI
Universidad Simón Bolívar, Caracas, Venezuela

SUMMARY. For Hypothesis Testing and Model Selection, the Bayesian approach is attracting considerable attention. The reasons for this attention include: i) it yields posterior probabilities of the models (and not simply accept-reject rules); ii) it is a predictive approach; and iii) it automatically incorporates the principle of scientific parsimony. Until recently, obtaining such benefits through the Bayesian approach required elicitation of proper subjective prior distributions, or the use of approximations (such as BIC) of questionable generality. In Berger and Pericchi (1996), the Intrinsic Bayes Factor Strategy was introduced, and shown to be an automatic default method corresponding to an actual (and sensible) Bayesian analysis. In particular, it was shown that the Intrinsic Bayes Factor yields an answer which is asymptotically equivalent to the use of a specific (and reasonable) proper prior distribution, called the Intrinsic Prior. Indeed, the IBF method can also be thought of as a method for constructing default proper priors appropriate for model comparisons. In this paper we study an implementation of the IBF strategy called the Median IBF. This seems to be a simple and very generally applicable IBF, which works well for nested or non-nested models, and even for small or moderate sample sizes; some of these situations can cause difficulties for other versions of IBFs.

1. **Introduction**

1.1 *Motivation and notation.* The following is a slight modification of a principle introduced in Berger and Pericchi (1996):

AMS (1991) *subject classification.* 62F15, 62F03, 62A15
Key words and phrases. Bayes factors; intrinsic Bayes factors; principle of model parsimony; posterior model probabilities.
* Research supported by the National Science Foundation(USA), Grant DMS-9303556, and by BID-CONICIT(Venezuela).

Principle 1:Testing and model selection methods should correspond, in some sense, to actual Bayes factors, arising from reasonable default prior distributions.

This principle provided the motivation for the development of Intrinsic Bayes Factors (IBFs). In Berger and Pericchi (1996), several versions of the IBF were introduced, including the Arithmetic (AIBF), Geometric (GIBF), Median, Trimmed, Expected, and Encompassing IBFs, but the primary emphasis was on the AIBF since it seemed best from the viewpoint of the above principle in the situation primarily studied, that of nested model comparisons. Scenarios in which the other IBFs are attractive were, however, discussed.

The reaction to the variety of IBFs that was presented was not entirely enthusiastic. Those who had been happy with use of BIC were put off by the additional complication of IBFs and, especially, with the need to use different IBFs in different scenarios. In addition, the papers of Bertolino and Racugno (1997) and O'Hagan (1997) demonstrated some of the difficulties that might be encountered in (casual) use of certain of the IBFs. These difficulties included the following:

1. The AIBF does not have multiple model coherence.

2. The AIBF and GIBF might be unstable with respect to small changes in the underlying improper priors or the data, particularly for small sample sizes.

3. The more complex model should be placed in the numerator of the AIBF. However, in some cases there is a difficulty in recognizing which is the more complex model, or all models have the same level of complexity.

4. Often, when comparing non-nested models, there is no natural encompassing model, i.e. , no natural model in which the models under study are nested. An encompassing model is required by the AIBF for non-nested models, except in situations where there is a group structure, such as in comparison of location-scale models.

5. The trimmed versions of the AIBF and GIBF have an extra parameter, α, indicating the degree of trimming that is chosen; this might be difficult to assess.

To address these concerns, we herein propose the Median IBF (already introduced in Berger and Pericchi, 1996), as a *single* implementation of the IBF strategy. We will put the Median IBF to the test in very different, and difficult, situations and show that it overcomes the aforementioned difficulties. While we do not feel that the Median IBF is the optimal IBF for each situation, it appears to be a good IBF in virtually any situation, and hence should have great appeal for those who want a single model selection strategy.

We begin with some needed notation. Suppose that we are comparing Models $M_1, M_2, ..., M_J$, for which the densities of the data $x = (x_1, \ldots, x_n)$ and the default priors for unknown parameters θ_j are $f_j(x|\theta_j)$ and $\pi_j^N(\theta_j)$, respectively, $j = 1, ..., J$. Then the marginal density of the observations under Model M_j and for the default prior is

$$m_j^N(x) = \int f_j(x|\theta_j)\pi_j^N(\theta_j)d\theta_j.$$

The (unscaled) Bayes Factor of M_j to M_i with respect to the default priors is then

$$B_{ji}^N(x) = \frac{m_j^N(x)}{m_i^N(x)}.$$

Denote by $x(l)$ any minimal training sample, i.e., a subset of the sample x such that $0 < m_j^N(x(l)) < \infty$ for all M_j, and no subset of it obeys that property. Similarly, denote by $x(-l)$ the remaining observations in x. For minimal training samples, the posterior distributions $\pi_j^N(\theta_j|x(l))$ are proper, and using these as "trained" priors results in (well scaled) Bayes Factors, based on the $x(-l)$,

$$B_{ji}(l) = \frac{m_j(x(-l)|x(l))}{m_i(x(-l)|x(l))},$$

where

$$m_k(x(-l)|x(l)) = \int f_k(x(-l)|\theta_k, x(l))\pi_k^N(\theta_k|x(l))d\theta_k, \; k = j, i.$$

Notice that, if all the factors are well defined, this simplifies to

$$B_{ji}(l) = B_{ji}^N(x)B_{ij}^N(x(l)). \qquad \ldots(1)$$

The Intrinsic Bayes Factors involve some form of averaging (with respect to l) of the $B_{ji}(l)$ or $m_k(x(-l)|x(l))$. The Arithmetic and Geometric IBF's are, respectively, the Arithmetic and Geometric averages of the $B_{ji}(l)$. In this article, we are going to study the following two versions of the Median IBF.

The first will be called the *Median IBF*, and is simply the natural

$$B_{ji}^M = MED[B_{ji}(l)],$$

where MED indicates the median. The second version that will be studied is the *Ratio of Medians IBF*, that is

$$B_{ji}^{RM} = \frac{MED[m_j(x(-l)|x(l))]}{MED[m_i(x(-l)|x(l))]}.$$

Both versions automatically obey the coherency requirement that $B_{ij}^M = 1/B_{ji}^M$, but only the second version is automatically coherent across two or more models. This, and that it is somewhat simpler when comparing several models, are the main strengths of B^{RM}. Its main disadvantage arises from the fact that it might not be a Bayes Factor from a real training sample (as is B^M), since the medians in the numerator and denominator are not necessarily obtained from the same training sample. As a consequence, B_{ji}^{RM} is not necessarily invariant with respect to monotonic transformations of x. Neither of these problems with B^M or B^{RM} are likely to be significant in practice, since training samples at which the medians are obtained will often be reasonably similar across models. Also, we will see that the two versions are typically quite close in value.

Before beginning this study, it is useful to give a general overview of the different default Bayes Factors (BF's) that have been proposed, together with a useful classification of BF's.

1.2 *Resampling and Sampling Bayes Factors.* A convenient taxonomy of Bayes Factors is into the two classes *Sampling BF's* and *Resampling BF's.* In what follows we briefly enumerate some of the BF's that lie in each class and some of their properties.

Sampling Bayes Factors: To this class belong the Bayes Factors that depend only on minimal sufficient statistics, and that do not divide the original sample into sub-samples (such as training samples or discrimination samples). To this group belong (among others): 1) Jeffreys' Conventional Prior BF's; 2) BIC, 3) Smith and Spiegelhalter's Global Bayes Factors, 4) Fractional BFs, and 5) Expected Intrinsic BF's and Bayes Factors using Intrinsic Priors. Sampling BF's are simpler computationally and tend to obey accepted principles such as the Likelihood Principle and the Sufficiency Principle.

Resampling Bayes Factors. This group is formed from the BF's that utilize training samples from the original sample, so as to allow use of improper default prior distributions. To this class belong (among others): 1) The original Intrinsic BF's, i.e., Arithmetic, Geometric and Median IBF's, and 2) Bayesian Cross Validation procedures, which typically employ maximal training samples.

Resampling BF's do not obey the Sufficiency Principle (since training samples are rarely functions of the sufficient statistic) and are typically computationally more intensive than sampling BF's. They tend to more closely follow Principle 1, however; see Berger and Pericchi (1997b) for some indications of this. Furthermore, there are indications that Resampling BF's might better adapt to unforeseen structures of the data, such as those arising from a misspecification of the candidate models. For one example of this, see Example 6 in Berger and Pericchi (1997b). Also, see Key, Pericchi and Smith (1997) for developments, applications and further motivations of resampling BF's, therein called "Global, Local and Comprehensive Divergences." This motivates our interest in exploring a default BF which is both a resampling BF and is not overly sensitive to small perturbations in the data or default priors.

2. Putting the Median IBF to the Test

2.1 *Example 1: Separate models of equal dimension for discrete data. Poisson versus Negative Binomial.* Suppose that a random sample has been generated from one of the following models:

$$M_1 : Po(x|\lambda) = \exp(-\lambda)\frac{\lambda^x}{x!},\ \ \lambda > 0.$$

$$M_2 : NegBin(x|1, \theta) = \theta(1-\theta)^x,\ \text{with } 0 < \theta < 1.$$

Now consider the family of priors, respectively, for the above two models:

$$\pi_1(\lambda) = \lambda^{-\alpha}\ \ \text{and}\ \ \pi_2(\theta) = \theta^{-\beta}(1-\theta)^{-\gamma},$$

where all α, β and γ belong to the interval $[0, 1]$. This example appears in Bertolino and Racugno (1997), henceforth denoted as BR (1997). Incidentally, this is one of Cox's examples in his original classical paper on Testing Separate Hypotheses.

There are many points to this example, but the most important are: i) These are models of equal dimension, and thus it is not clear which should be placed in the numerator of the Arithmetic IBF, and ii) The data is discrete and thus the number of training samples might change dramatically for different improper priors.

BR (1997) considers two different (small) samples which share the same Maximum Likelihood Estimators under both models:

$$x^{(1)} = (1, 1, 1, 1, 1, 1, 1, 1, 2, 2)\quad \text{and}\quad x^{(2)} = (0, 0, 0, 0, 1, 1, 2, 2, 3, 3).$$

Furthermore, BR (1997) considers quite different values of the prior hyperparameters, namely ($\alpha = \beta = 1$, $\gamma = 0.5$), ($\alpha = 1$, $\beta = \gamma = 0.5$), ($\alpha = \beta = \gamma = 1$), and ($\alpha = \beta = \gamma = 0$). (Note that we would have recommended using "reference priors"; here that would correspond to the "Jeffreys priors", being given by ($\alpha = \gamma = 0.5$, $\beta = 1$). This would have eliminated most of the reported problems.) Also considered in BR (1997) were slight variations of α, say from 1 to 0.99, which change the training samples in a fundamental manner, since a 0 observation cannot be a training sample when $\alpha = 1$, but can be a training sample when $\alpha = 0.99$. In Tables 1 and 2 we present the Arithmetic IBF and the Median IBF for the various priors and two samples, respectively.

Table 1. ARITHMETIC AND MEDIAN IBFs FOR THE FIRST SAMPLE.

α	β	γ	$B_{12}^A(x^{(1)})$	$B_{12}^M(x^{(1)})$	$B_{21}^A(x^{(1)})$	$B_{21}^M(x^{(1)})$
1	1	0.5	33.200	31.95	0.0303	0.0313
0.99	1	0.5	34.575	32.15	0.0293	0.0311
1	1	1	35.198	35.21	0.0284	0.0284
0.99	1	1	35.379	35.46	0.0283	0.0282
1	0.5	0.5	28.403	28.41	0.0352	0.0352
0.99	0.5	0.5	28.548	28.57	0.035	0.035
0	0	0	26.715	29.67	0.0404	0.0337

Table 2. ARITHMETIC AND MEDIAN IBFs FOR THE SECOND SAMPLE.

α	β	γ	$B_{12}^A(x^{(2)})$	$B_{12}^M(x^{(2)})$	$B_{21}^A(x^{(2)})$	$B_{21}^M(x^{(2)})$
1	1	0.5	1.1315	1.06	0.981	0.9398
0.99	1	0.5	0.6278	0.89	15.519	1.1197
1	1	1	0.9777	0.98	1.0228	1.0228
0.99	1	1	0.9766	0.97	1.0240	1.0257
1	0.5	0.5	0.7725	0.79	1.2956	1.2675
0.99	0.5	0.5	0.4758	0.79	13.364	1.2585
0	0	0	1.2863	0.82	1.6979	1.2128

The main messages of this example concerning the Arithmetic IBF are that, for the troublesome sample $x^{(2)}$: i) the Arithmetic IBF is quite sensitive to "slight" variations of the improper prior; and ii) it does matter which of the two models is placed in the numerator of the Arithmetic IBF. For the Median IBF, however, neither of these two difficulties arises. The Median IBF is extremely insensitive to variation of the improper priors and is, of course, coherent in the sense that $B_{21}^M = 1/B_{12}^M$. Also, for the first sample, for which the Arithmetic IBF behaves very regularly, the Median IBF is quite close to the Arithmetic IBF. Indeed, in this example the Median IBF deviates from the Arithmetic IBF only if the latter shows irregular behavior. We do not separately present results for the Median IBF and Ratio of Medians IBF, since they coincide exactly in this example.

2.2 *Example 2: Only one informative observation.* There is a group of examples with the theme that only one of the observations is informative for both the discrimination of models and estimation of parameters. We analyze two of these examples.

EXAMPLE 2.2.1: *Nested hypotheses for discrete data.* This example appears in O'Hagan (1997). It involves a hypothesis test of a Bernoulli parameter $\theta = P(X = 1|\theta)$. It is desired to test

$$M_0 : \theta = \theta_0 \text{ versus } M_1 : \theta \neq \theta_0.$$

We will be concerned below with $\theta_0 = 0$, in which case any nonzero Bernoulli observation clearly establishes that M_1 is true.

Suppose, now, that the improper prior $\pi^H(\theta) \propto \theta^{-1}(1-\theta)^{-1}$ is to be used with some IBF, but that the Bernoulli observations consist of all zeroes except for one 1. Then, it is easy to see that any minimal training sample must contain exactly one 0 and one 1, and thus the remaining data $x(-l)$, which will be used to compute the Bayes factor corresponding to the prior based on the training sample, will contain only zeroes. No IBF will then conclude that M_1 is true for sure, as is desired when $\theta_0 = 0$.

The Median IBF (or any other version of the IBF) is less sensitive to the difficulty with π^H if at least two ones are observed. If r ones are observed, all training samples will still have one 0 and one 1, and all IBFs will then equal

$$B_{10}^I = \frac{\Gamma(r)\Gamma(n-r)}{\Gamma(n)\theta_0^r(1-\theta_0)^{n-r}}\theta_0(1-\theta_0).$$

This will be ∞ as long as $r > 1$ (when $\theta_0 = 0$), so that the IBFs will correctly conclude that M_1 is true.

It should also be mentioned that the improper prior π^H is quite unreasonable. Again, we always recommend use of the reference prior (here, also, the Jeffreys prior), which is $\pi^J(\theta) = \frac{1}{\pi}\theta^{-1/2}(1-\theta)^{-1/2}$; the Uniform prior is also quite reasonable for a Bernoulli parameter. Since these priors are proper, a minimal training sample would be empty, and so IBFs are just the original Bayes factors. For example, if the Jeffreys prior is assumed, and only one 1 is observed, then the Bayes Factor is

$$B_{10}^{\pi^J} = \frac{\Gamma(n-1/2)}{\Gamma(1/2)\Gamma(n)\theta_0(1-\theta_0)^{n-1}} ,$$

which is ∞ for $\theta_0 = 0$, as it should be.

EXAMPLE 2.2.2: *ested hypotheses with continuous data.* This is a simplified version of an example that also appears in O'Hagan (1997). Suppose that M_1 states that the X_i are a random sample from a Normal Distribution with mean 0 and known variance σ_0^2, while M_2 is the same except that observation X_j is distributed as a Normal but with mean shifted by an unknown δ. Since X_j is the only informative data about the parameter δ, it will be contained in any training sample, and thus the *Resampling* IBF's will all equal one (no matter how large the 'outlier').

It is indeed clear that one cannot use a resampling IBF if there is only one relevant sample. However, one can still use the intrinsic Bayes factor approach to determine Intrinsic Priors, which can then be used directly. The resulting procedure will be to reject M_1 in favor of the outlier model when the standardized residual of x_i is greater than a threshold value; this agrees with standard practice and has the added attraction that the value of the threshold is given in terms of (easy to interpret) Bayes factors (or posterior probabilities) for the intrinsic prior. Note that derivation of the intrinsic prior is essentially based on 'imagining' a very large sample (of 'outliers') and deriving the intrinsic prior asymptotically; hence the size of the actual sample becomes irrelevant.

2.3 Example 3: *Separate models of different dimensions for continuous data.* This example also appears in Bertolino and Racugno (1997). We compare the Exponential versus the LogNormal model (an example analyzed in Berger and Pericchi, 1996, with different data), using the usual reference priors and with the following small samples:

$$x^{(1)} = (0.7, 1, 1, 1, 1, 1, 2, 3, 4, 5), \quad x^{(2)} = (0.7, 0.999, 0.9999, 1, 1.0001, 1.001, 2, 3, 4, 5).$$

The second sample is only a slight variant of the first ("a more accurate experiment"), but the difference can have a significant effect because the number of proper training samples for each of the samples is quite different. This happens because a minimal training sample is any pair of different observations, and the first sample has 5 equal observations, while the second sample has none. As a consequence, for the second sample there are 45 training samples, but for the first sample 10 of these are not training samples. (The "near singularity" of these 10 training samples for the second data set is also a potential problem.) Avoiding these potential difficulties was the reason that Berger and Pericchi (1996) proposed use of α-trimmed and Median IBFs.

In BR (1997), the Arithmetic IBFs are computed, separately, as B_{21}^A and B_{12}^A, although it is natural to think that M_2 is more complex than M_1; indeed, Berger and Pericchi (1997a) show that this is so, in the sense that the expectation of the corrections for B_{21}^A are finite, while this is not so for B_{12}^A. Hence the prescription in Berger and Pericchi (1996) of *defining* B_{12}^A as the inverse of B_{21}^A should have been used.

The results for this example are displayed in Table 3.

Table 3. ARITHMETIC AND MEDIAN IBFS FOR EXPONENTIAL VERSUS LOGNORMAL.

Sample	B^A_{12}	B^M_{12}	B^{RM}_{12}	B^A_{21}	B^M_{21}	B^{RM}_{21}
$x^{(1)}$	0.8475	0.6568	0.688	1.3929	1.5225	1.4534
$x^{(2)}$	345.71	0.7974	0.643	1.0837	1.2541	1.5559

Table 3 is revealing, and suggests the following comments:

1. The pathological behavior of B^A_{12}, for $x^{(2)}$, is evident but, again, these should have been calculated as $B^A_{12} = 1/B^A_{21}$. Thus the correct values of B^A_{12} for the first and second samples, respectively, are 0.7179 and 0.9228, instead of 0.8475 and 345.71. The crazy behavior vanishes if the correct definition of the Arithmetic IBF is applied.

2. Nevertheless, the observation in BR (1997), that there are cases when there is no clear "more complex model", is important to note. For an automatic implementation of the IBF strategy, it is useful to have a tool which does not depend on which model is more complex, and that is coherent, in the sense that $B_{21} = 1/B_{12}$. The Median IBF, in its two version, obeys these requirements, and automatically avoids the pathological behavior of the direct calculation of B^A_{12} for the second sample.

3. The two versions of the Median IBF are quite stable for the two samples and are also quite close to each other.

2.4 EXAMPLE 4: *Separate scale models and different improper priors.* This is the final example in BR (1997). Here there is a comparison between three separate scale models:

$$M_1 : Normal, \ \ M_2 : Laplace \ \text{and} \ M_3 : Cauchy,$$

all with location assumed known. Let us denote, respectively, by σ, ρ and τ the scale parameters under each of the three models. BR (1997) considers the following family of improper priors, indexed by the hyperparameters α, β and γ:

$$\pi_1(\sigma) = \sigma^{(2\alpha-3)}, \quad \pi_2(\rho) = \rho^{(\beta-2)}, \quad \text{and} \quad \pi_3(\tau) = \tau^{(1-2\gamma)}.$$

This is an example of continuous separate (i.e., non-nested) models of equal complexity, and so it is not clear which to place in the numerator of the Arithmetic IBF. BR (1997) chooses different values of the hyperparameters to show that the Arithmetic IBF changes with the priors. Also, for certain values of the hyperparameters, the Arithmetic IBF can be incoherent, in the sense that $B^A_{12} > 1$ and $B^A_{21} > 1$ simultaneously.

Note, first, that, for the very wide class of problems on which a group structure is acting, e.g. location or location-scale models, there is a very strong reason to use the reference prior; see Berger and Pericchi (1996, 1997a,1997b) and Berger, Pericchi and Varshavsky (1996). Indeed, a remarkable simplification then occurs: for scale parameter models, this simplification (noting that a minimal training sample is a single observation) is that

$$m_j^N(x_l) = \frac{1}{2|x_l - \mu_0|}, \quad j = 1, 2, 3. \qquad \ldots(2)$$

This fact implies an enormous simplification in IBFs, since then all the training sample 'corrections' cancel out. This also establishes that the reference priors are "well calibrated" for this problem. Thus, using reference priors for the problem implies that all versions of IBFs coincide with the Bayes Factor with respect to the non-informative priors, i.e.,

$$B_{ij}^I = B_{ij}^N. \qquad \ldots(3)$$

For non reference priors, relationships (2) and (3) are no longer true.

In spite of our strong preference for reference priors in this problem, we put the Median IBF to the test for other improper priors as well. The reference priors, proportional to the above prior with ($\alpha = \beta = \gamma = 1$) in the given parameterizations, are considered, as well as the perturbations ($\alpha = 1,\ \beta = 0.5,\ \gamma = 1.25$) and ($\alpha = 1,\ \beta = 0.75,\ \gamma = 0.75$). We only present this subset of the values considered in BR (1997), since the behavior for other values is similar.

As seen from (3), the Arithmetic and Median IBF coincide for the reference priors, but they differ quite markedly for other priors. We do not think that the Arithmetic IBF should be used for separate models without encompassing or without a matching relationship such as (2), but we present the corresponding numbers in the next Table for sake of comparison. Table 4 considers the same set of data used in BR (1997), namely $x_i - \mu_0 = (-1, -0.4, -0.2, 0.001, 0.01, 0.1, 0.3, 1)$. Only the Median IBF is presented, because the Ratio of Medians IBF was quite similar. For instance, its corresponding values for the first non-reference prior, i.e., the second row in the table, and following the order of the Median IBF in the table, were $2.31, 1.48$ and 0.64. Notice, also, that the values obtained for the Median IBF are almost exactly coherent across the three models. (Of course, the Ratio of Medians IBF would achieve this automatically.)

Table 4. MEDIAN AND ARITHMETIC IBF FOR SEPARATE SCALE MODELS, WITH REFERENCE AND NON REFERENCE PRIORS.

α	β	γ	B_{21}^M	B_{21}^A	B_{12}^A	B_{31}^M	B_{31}^A	B_{13}^A	B_{32}^M	B_{32}^A	B_{23}^A
1	1	1	2.64	2.64	0.38	1.88	1.88	0.53	0.71	0.71	1.41
1	0.5	1.25	2.38	2.44	1.37	1.52	1.56	2.14	0.64	0.64	1.57
1	0.75	0.75	2.59	2.42	0.56	1.27	4.09	0.82	0.49	4.76	2.53

As anticipated, the Arithmetic IBF is unstable for the last non-reference prior (third row) and, for both non-reference priors, the Arithmetic IBF and its reciprocal might be simultaneously bigger than one. (We do not show the reciprocal of the Median IBF, since $B_{ij}^M = 1/B_{ji}^M$ automatically). On the other hand, the Median IBF is extremely stable and the same ranking of models is obtained for both priors.

2.5 Example 5: *Separate regressions.* In this example, we study Hald's Regression Data, typically analyzed using normal regression models. There are four potential regressors, denoted by 1,2,3,4, and a constant included in all models and denoted by c. Let us compare the following non nested models:

$$M_2 : \{c, 1, 2\} \text{ versus } M_4 : \{c, 1, 4\} .$$

Berger and Pericchi (1996) recommended the encompassing approach, whereby these two models are compared first against the biggest possible model M_0 : $\{c, 1, 2, 3, 4\}$, computing the Arithmetic IBFs B_{0j}^{0A} for j=2 and 4, and then forming the encompassing Arithmetic IBF by

$$B_{24}^{0A} = B_{04}^{0A} / B_{02}^{0A},$$

with the reciprocal of this expression being B_{42}^{0A}. In order to do this, the training sample sizes need to be large enough to accommodate the encompassing model. Here, for instance, the minimal training sample size, using reference priors and the encompassing approach, are 6; a direct comparison between M_2 and M_4 would only require a training sample of size 4.

The goal here is to compare the Median IBF, for a direct comparison of M_2 and M_4 , with the Arithmetic and Geometric IBFs under the encompassing approach. We use reference priors, i.e., the reciprocal of the standard deviation. For the Geometric IBF, we used a training sample of size 6 also, viewing it as an 'encompassing' Geometric IBF. For the Median IBF, however, we only used a training sample size of 4, since it is a direct comparison. The answers obtained via all approaches are quite similar, especially considering that there was a very small sample size and a high degree of multicolinearity between the covariates of the two models.

Table 5: HALD'S DATA; SEPARATE REGRESSIONS; MEDIAN AND ENCOMPASSING IBF'S.

B_{24}^{MR}	B_{24}^{RM}	B_{24}^{0A}	B_{24}^{0G}
3.78	2.87	2.55	2.5

2.6 *Example 6: Nested regressions.* We again consider Hald's data from the previous example, but now do nested comparisons of the full model M_0 to M_2

and to M_4 ; the minimal training sample size is again 6. We consider the effects of two (very) different priors, the reference prior and the Jeffreys prior, denoted by π^r and π^j, respectively.

Table 6. HALD'S DATA; NESTED REGRESSIONS; MEDIAN, GEOMETRIC AND ARITHMETIC IBF'S.

Prior	B_{02}^A	B_{02}^G	B_{02}^M	B_{02}^{RM}	B_{04}^A	B_{04}^G	B_{04}^M	B_{04}^{RM}	B_{24}^A	B_{24}^G	B_{24}^M	B_{24}^{RM}
π^r	.18	0.08	.20	.17	.46	0.20	.47	.44	2.55	2.5	2.32	2.60
π^j	.16	0.004	.026	.016	.41	0.01	.09	.04	2.55	2.49	3.40	2.73

For the reference prior, which we recommend, the Median IBFs are quite close to the Arithmetic IBF. However, the (encompassing) Arithmetic IBF is clearly the most stable with respect to change in the prior. This, in part, is why we feel that the Arithmetic IBF is most suitable for nested comparisons, at least when comparing just two nested models. For more complex comparisons that involve several models, and for non-nested comparisons, the Median IBF has a clear advantage, being very safe to use, especially when coupled with reference priors. We do not recommend the Median IBF for nested comparisons with the Jeffreys prior. Note that the Geometric IBF seems to be too small for all priors, but particularly for the Jeffreys prior.

2.7 EXAMPLE 7: *Proschan data.* Here we reanalyze the Proschan Data, which consists of 30 failure times of the air conditioning system of an airplane. As in BP (1996), we compare the models

$$M_1 : Exponential, \quad M_2 : LogNormal, \quad \text{and} \quad M_3 : Weibull,$$

using both the reference priors (denoted by π^r) and Jeffreys priors (denoted by π^j).

Let Aij, Gij and Mij, denote the Arithmetic, Geometric and Median IBF, respectively, for Model i over Model j. The primary goal here is to study the coherence, across models, of the Median IBF. (We do not give the Ratio of Medians IBF because it is automatically coherent across models.)

Table 7: PROSCHAN DATA, MEDIAN, GEOMETRIC AND ARITHMETIC IBF.

Prior	A21	G21	M21	A31	G31	M31	A32	G32	M32	A23	G23	M23
π^r	.37	.33	.42	.26	.23	.29	.7	.7	.7	1.42	1.42	1.42
π^j	.37	.33	.42	.25	.15	.29	.66	.46	.61	3.93	2.15	1.65

From Table 7, it can be concluded that the Median IBF is less sensitive to changes in the prior than the other IBFs. Also, it is very close to being completely coherent across models. For example, we can compute Mij directly, or we can calculate it via Mkj/Mki, which we call M*ij. Table 8 presents the results of these two ways of computing the Bayes factors.

Table 8: ALMOST COHERENCE OF THE MEDIAN IBF FOR THREE MODELS.

Prior	M31	M*31	M32	M*32	M21	M*21
π^r	.29	.30	.7	.69	.42	.41
π^j	.29	.25	.61	.69	.42	.47

The Median IBF is almost exactly coherent, especially for reference priors. The Geometric IBF also has good behavior here, except that it is slightly more sensitive than the Median to changes in the priors. The same could be said about the Arithmetic IBF, except that the behavior of B^A_{23} with the Jeffreys prior is quite different than that of the reciprocal of B^A_{32}.

3. **Intrinsic priors**

The extent to which the Median IBFs satisfy Principle 1 is clearly of interest. Intrinsic priors can exist for either, as the following two examples show.

3.1 *Location-Exponential testing.* Suppose $X_1, X_2, ..., X_n$ are i.i.d. with density

$$f(x_i|\theta) = \exp[-(x_i - \theta)]\ I_{(\theta,\infty)}(x_i),$$

where "I" denotes the indicator function. It is desired to compare

$$M_1 : \theta = \theta_0 \ \text{ versus } \ M_2 : \theta \neq \theta_0,$$

employing the usual non-informative prior $\pi_2^N(\theta) = 1$. Computation yields

$$m_1^N(x_1, x_2, ..., x_n) = \exp(n\theta_0 - S) I_{(\theta_0,\infty)}(x_{min}),$$

$$m_2^N(x_1, x_2, ..., x_n) = \frac{1}{n}\exp(nx_{min} - S),$$

where $S = \sum_{i=1}^n x_i$ and $x_{min} = min[x_1, ..., x_n]$. Hence,

$$B_{21}^N(x) = \begin{cases} \frac{1}{n}\exp[n(x_{min} - \theta_0)] & \text{if } x_{min} > \theta_0 \\ \infty & \text{if } x_{min} < \theta_0. \end{cases}$$

Any single observation is a minimal training sample, and the resulting "trained priors" are

$$\pi_2^N(\theta|x_i) = \exp(\theta - x_i) I_{(-\infty,x_i)}(\theta). \qquad \ldots(4)$$

Using these to compute the IBF's yields, for the Arithmetic and Median versions,

$$B_{21}^A = \begin{cases} \frac{1}{n}\exp(n(x_{min} - \theta_0))[\frac{1}{n}\sum_{i=1}^n \exp(\theta_0 - x_i)] & \text{if } x_{min} > \theta_0 \\ \infty & \text{if } x_{min} < \theta_0 \end{cases}$$

$$B_{21}^M = \begin{cases} \frac{1}{n}\exp(n(x_{min} - \theta_0))\exp(\theta_0 - MED[x_i]) & \text{if } x_{min} > \theta_0 \\ \infty & \text{if } x_{min} < \theta_0. \end{cases}$$

(A technical point is in order: The simplifying formula (1) does not apply here due to the presence of terms of the form $\infty \cdot 0$; hence one must compute the IBF's using the "trained priors" in (4).)

As $n \to \infty$, the $\{\theta < \theta_0\}$ are irrelevant, since then $x_{min} < \theta_0$ (and the IBF's equal ∞) with probability one. As $n \to \infty$ for $\theta > \theta_0$,

$$\frac{1}{n}\sum_{i=1}^n \exp(\theta_0 - x_i) \to E_\theta[\exp(\theta_0 - X)] = \frac{1}{2}\exp[-(\theta - \theta_0)],$$

and also

$$\exp(\theta_0 - MED[x_i]) \to \exp(\theta_0 - [\theta + \log(2)]) = \frac{1}{2}\exp[-(\theta - \theta_0)].$$

Since $\pi^N(\theta) = 1$ and M_1 is nested in M_2, equation (49) in Berger and Pericchi (1996) shows that the intrinsic priors for both B_{21}^A and B_{21}^M are

$$\pi_2^I(\theta) = \begin{cases} arbitrary & \theta < \theta_0 \\ \frac{1}{2}\exp[-(\theta - \theta_0)] & \theta > \theta_0. \end{cases}$$

Note that $\int_{\theta_0}^\infty \pi_2^I(\theta)d\theta = \frac{1}{2}$, so that proper intrinsic priors clearly exist which assign equal mass to $\{\theta < \theta_0\}$ and $\{\theta > \theta_0\}$.

3.2 *Scale parameter problems.* Suppose $X_1, ..., X_n$ are i.i.d. with density of the form

$$f(x_i|\theta) = \frac{1}{\theta} g(\frac{x_i}{\theta}),$$

for $x_i > 0$ and $\theta > 0$, where $g(.)$ is monotonic. It is desired to compare

$$M_1 : \theta = \theta_0 \ \text{ versus } \ M_2 : \theta \neq \theta_0,$$

utilizing the non-informative prior $\pi_2^N(\theta) = 1/\theta$. Clearly a single observation, x_i, is a minimal training sample, and

$$B_{12}^N(x_i) = \frac{1}{\theta_0} g(\frac{x_i}{\theta_0}) / \int_0^\infty \frac{1}{\theta^2} g(\frac{x_i}{\theta}) d\theta,$$

which can be proved to be equal to

$$B_{12}^N(x_i) = \frac{1}{\theta_0} g(\frac{x_i}{\theta_0})/(\frac{1}{x_i}),$$

from which expressions for the various IBF's follow directly.

It is somewhat difficult to study intrinsic priors for the Median IBF, but the Ratio of Medians IBF is

$$B_{21}^{RM} = B_{21}^N \cdot \frac{MED[\theta_0^{-1} g(x_i/\theta_0)]}{MED[1/x_i]},$$

which is equal to

$$B_{21}^N \cdot \frac{\theta_0^{-1} g(MED[x_i]/\theta_0)}{1/MED[x_i]},$$

using the monotonicity of $g(\cdot)$. As $n \to \infty$, it can be proved that $MED[x_i] \to \theta\ m$, where m is the median of f when $\theta = 1$. Hence, using expression (49) in Berger and Pericchi (1996), the intrinsic prior is determined as

$$\pi_2^I(\theta) = \frac{\theta_0^{-1} g(\theta m/\theta_0)}{1/(\theta m)} \cdot \frac{1}{\theta} = \frac{m}{\theta_0} \cdot g(\frac{\theta}{\theta_0/m}),$$

which is clearly proper. Furthermore, the prior probability that $\{\theta < \theta_0\}$ equals 0.5, so that the intrinsic prior is admirably "balanced".

As a specific illustration, consider the usual Exponential model, $f(x_i|\theta) = \theta^{-1}\exp(-x_i/\theta)$. It is desired to test $M_1 : \theta = \theta_0$ versus $M_2 : \theta \neq \theta_0$ and the nonnformative prior $\pi^N(\theta) = 1/\theta$ is employed. Then the Ratio of Medians Intrinsic prior is

$$\pi_2^I(\theta) = \frac{\log 2}{\theta_0} \exp(-\theta \log 2/\theta_0),$$

an Exponential prior with mean parameter $\theta_0/\log 2$.

It is unfortunately, rather rare for the Median IBFs to have a proper intrinsic prior; more typical is the situation in the following example.

3.3 *Multivariate normal testing.* Suppose $X_1, ..., X_n$ are i.i.d. from the p-variate normal distribution with mean vector θ and identity covariance matrix. It is desired to compare

$$M_1 : \theta = 0 \quad \text{versus} \quad M_2 : \theta \neq 0.$$

Utilizing the standard noninformative prior, $\pi_2^N(\theta) = 1$, a minimal training sample is a single vector x_i, and

$$B_{12}^N(x_i) = (2\pi)^{-p/2}\exp(-|x_i|^2/2).$$

The Arithmetic IBF and Median IBF are

$$B_{21}^A = B_{21}^N \cdot (2\pi)^{-p/2}\frac{1}{n}\sum_{i=1}^{n}\exp(-|x_i|^2/2),$$

$$B_{21}^M = B_{21}^N \cdot (2\pi)^{-p/2}\exp(-MED[|x_i|^2]/2).$$

As $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\exp(-|x_i|^2/2) \to E_\theta[\exp(-|x_i|^2/2)] = 2^{-p/2}\exp(-|\theta|^2/4),$$

$$MED[|x_i|^2] \to p + \frac{1}{6}(\frac{1}{p} - 5) + |\theta|^2(1 - \frac{1}{6p}) \text{ (approximately).}$$

(The last expression is, for our purposes, a reasonably accurate approximation to the median of the corresponding non-central chi square distribution.) It is straightforward to see that the resulting intrinsic prior for the Arithmetic IBF, $\pi_2^A(\theta)$, is a $N_p(0, 2I)$ density, while that for the Median IBF is

$$\pi_2^M(\theta) = \exp(-\frac{1}{2}[p + \frac{1}{6}(\frac{1}{p} - 5)]) \cdot (2\pi)^{-p/2}\exp(-\frac{1}{2}(1 - \frac{1}{6p})|\theta|^2).$$

This last prior is not proper, integrating to approximately $K_p = (1.09)\exp(-\frac{1}{2}[p + \frac{1}{6}(\frac{1}{p} - 5)])$, values of which are given in Table 9 for various p. Also given in the Table are the ratios of the intrinsic priors evaluated at zero, for the Median IBF and the Arithmetic IBF.

Table 9.

p	1	2	3	4	5	10
K_p	0.93	0.58	0.36	0.22	0.13	0.01
$\pi_2^M(0)/\pi_2^A(0)$	1.20	1.07	0.93	0.80	0.69	0.32

The numbers in Table 9 accurately reflect our experience. For comparing models of similar dimension, the intrinsic priors for the Median IBF fail to be proper only by a moderate constant but, for large dimensional differences, the constants can be much smaller than one. This discrepancy is tempered, however, by the fact that $\pi_2^M(\theta)$ will typically have a smaller spread than will $\pi_2^A(\theta)$. One consequence is that the values of the two priors at zero are quite similar for moderate p, and the values at zero are often the most influential features of the priors in determination of the Bayes Factor. This probably explains why, in application to nested models, the Arithmetic IBF and Median IBF tend to give quite similar answers, at least for reference priors.

4. Conclusion

For comparison of two nested models with a moderate amount of data, we would still recommend use of the Arithmetic IBF (or an expected version), because of its guaranteed correspondence with a (sensible) proper intrinsic prior. If, however, the data set is small, or a single default Bayes factor methodology is desired, we would recommend the Median IBF. The amount by which the Median IBF "violates" Principle 1 appears to be quite modest in practice, and its enormous range of applicability and stability with respect to small training samples would argue for its general use. This is especially so if it is combined with reference priors.

Acknowledgments. This research was supported by the National Science Foundation (USA), Grant DMS-9303556, and by BID-CONICIT (Venezuela). The authors are grateful to F. Bertolino and W. Racugno for their paper, which provided the motivation for this more systematic study of the Median IBF. N. Nadal and J. M. Pérez kindly performed some of the calculations for us.

References

Berger, J. and Mortera, J. (1997). Default Bayes factors for one-sided hypothesis testing, *ISDS Discussion Paper* 97-44, Duke University.

Berger, J. and Pericchi, L. (1995). The intrinsic Bayes factor for linear models, in *Bayesian Statistics* V, eds. J. M. Bernardo, *et. al.*, London: Oxford University Press, pp. 23–42.

− − −− (1996). The intrinsic Bayes factor for model selection and prediction, *Jour. Amer. Statis. Assoc.*, **91**, 109–122.

− − −− (1997a). On the justification of default and intrinsic Bayes factors, in *Modeling and Prediction*, eds. J. C. Lee, *et. al.*, New York: Springer-Verlag, pp. 276–293.

− − − (1997b). On criticisms and comparisons of default Bayes factors for model selection and hypothesis testing, *ISDS Discussion Paper* 97-43, Duke University.

Berger, J., Pericchi, L., and Varshavsky, J. (1996). Bayes factors and marginal distributions in invariant situations, *Technical Report* 95-7C, Purdue University, Department of Statistics.

Bertolino, F. and Racugno, W. (1997). Is the intrinsic Bayes factor intrinsic?, *Metron* **LIV**, 5–15.

Dmochowski, J. (1994). Intrinsic priors via Kullback-Liebler geometry, *Technical Report* 94-15, Purdue University, Department of Statistics.

Key, J. T., Pericchi, L. R., and Smith, A. F. M. (1997). Choosing among models when none of them are true, *Technical Report*, Imperial College London.

O'Hagan, A. (1995). Fractional Bayes factors for model comparisons, *Jour. Royal Statis. Soc.*, Ser. B, 57, 99–138.

– – –– (1997). Properties of intrinsic and fractional Bayes factors, *Test* **6**, 101–118.

J.O. Berger
ISDS
Duke University
Box 90251
Durham, NC 27708-0251
USA
e-mail : berger@stat.duke.edu

L.R. Pericchi
CESMa
Universidad Simón Bolívar
Apartado 89000
Caracas 1080-A
Venezuela
e-mail : pericchi@usb.ve