Bayesian Robustness
IMS Lecture Notes - Monograph Series (1996) Volume 29

ON THE ROBUSTNESS OF THE INTRINSIC BAYES FACTOR FOR NESTED MODELS [1]

BY BRUNO SANSÓ, LUIS R. PERICCHI AND ELÍAS MORENO
Universidad Simón Bolívar and Universidad de Granada

Bayesian model comparisons are known to be undetermined when improper priors are employed. The Intrinsic Bayes factor (IBF) is a general automatic procedure for model comparison proposed in Berger and Pericchi (1993) which addresses the difficulties that arise when improper priors are employed. An appealing justification of the IBF is that it asymptotically corresponds to actual Bayes factors of particular priors. Such priors are called instrinsic priors and can be obtained as solutions of two functional equations. In this paper we consider issues related to the robustness of the IBF in the nested model situation.

1. Introduction. The problem of comparing two models is addressed, from a Bayesian perspective, in the following way: consider a set of data X that have density $f_i(\mathbf{x}|\theta_i)$ under model $M_i, i = 1, 2, \theta_i \in \Re^{k_i}$, and suppose that prior distirbutions $\pi_i(\theta_i)$ are selected for the parameters of each model. Select prior probabilities for each model and update them using the Bayes factor defined as

$$B_{21} = \frac{\int f_2(\mathbf{x}|\theta_2)\pi_2(\theta_2)d\theta_2}{\int f_1(\mathbf{x}|\theta_1)\pi_1(\theta_1)d\theta_1} \quad . \tag{1}$$

The same tool is not available for comparing models with improper priors, since these are defined only up to a multiplicative constant, leaving the Bayes factor undetermined. Berger and Pericchi (1995) introduced the idea of *intrinsic Bayes factor* (IBF) to address this problem. The method consists of using a subsample of minimal size to obtain a proper posterior distribution that is used as a proper prior to compute a Bayes factor for the remainder of the data, the results are averaged over all possible training samples to produce the IBF. More precisely, letting L be the number of all training samples, the arithmetic version of the IBF is defined as

$$B_{21}^{AI} = B_{21}^{N} \frac{1}{L} \sum_{l=1}^{L} B_{12}^{N}(\mathbf{x}(l)) \tag{2}$$

where $B_{21}^N = m_2^N(\mathbf{x})/m_1^N(\mathbf{x})$; $B_{12}^N(\mathbf{x}(l)) = m_1^N(\mathbf{x}(l))/m_2^N(\mathbf{x}(l))$,

$$m_i^N(t) = \int f_i(t|\theta_i)\pi_i^N(\theta_i)d\theta_i \quad ,$$

[1]This research was partially supported by BID-CONICT number I-06, Venezuela and DIGYCIT number PB93-11154, Spain.
AMS 1991 subject classifications. Primary 62F15; secondary 62F35.
Key words and phrases. Bayesian model comparison. Bayesian robustness. Bayes factors. intrinsic priors

π_i^N denotes the noninformative prior of the ith model, and $\mathbf{x}(l)$ is a minimum training sample, that is, a subsample of minimal size such that $0 < m_i^N(\mathbf{x}(l)) < \infty \;\; i = 1, 2$ and no subset of it satisfies the condition.

In Berger and Pericchi (1995) the authors later found that, under certain regularity conditions, the IBF is asymptotically equal to the actual Bayes factor for a particular prior distribution, which they coined an *intrinsic prior* for the problem. Intrinsic priors, and hence their Bayes factors, are not uniquely determined, in the second section of this paper we explore the robustness of the Bayes factors for nested models as the prior distributions are varied over the class of intrinsic prior distributions.

In the third section we consider the case where the Bayes factor is seen to be highly sensitive to the choice of the proper prior under one of the models, a situation pointed out in Berger (1985) and references therein, and we address the problem using an IBF approach. In the last section of this paper we consider how the choice of the improper prior for the scale parameter of a normal linear model affects the resulting IBF and intrinsic priors. Finally we present some conclusions and open problems.

An alternative solution to the indeterminacy of the Bayes factor, when improper priors are used, is the *fractional Bayes factor* (FBF) proposed in O'Hagan (1995): in De Santis and Spezzaferri (1995) some robustness properties of the FBF are studied. However it is unclear if it exists for the FBF a class similar to the class of intrinsic priors.

2. The class of intrinsic priors. Given default priors $\pi_i^N(\theta_i)$ as in (2), the *class of intrinsic priors* is given by those measures $\pi_i^I(\theta_i)$ for which the corresponding Bayes factor B_{21} tends to B_{21}^{AI} as the sample size n tends to infinity. This is a procedure for automatic assessment of a class of priors, a difficult issue in robust Bayes.

The interest of studying such a class has several layers. In the first place the properties of such a class are to be described and we can judge if the class appears to be reasonable. In the second place understanding the class sheds light on the conditions for which the asymptotics are valid. In the third place, the extent of the robustness achieved within the class has to be evaluated for finite sample sizes.

Letting $\hat{\theta_1}$ and $\hat{\theta_2}$ denote the MLE of f_1 and f_2 respectively, under appropiate regularity conditions for the likelihoods to concentrate around the MLEs, and for continuous priors π_1 and π_2, the Laplace expansion of (1) yields

$$B_{21} = B_{21}^N \frac{\pi_2(\hat{\theta_2})\pi_1^N(\hat{\theta_1})}{\pi_2^N(\hat{\theta_2})\pi_1(\hat{\theta_1})}(1 + o(1)) \quad . \tag{3}$$

Equating (2) and (3) gives

$$\frac{\pi_2(\hat{\theta_2})\pi_1^N(\hat{\theta_1})}{\pi_2^N(\hat{\theta_2})\pi_1(\hat{\theta_1})}(1+o(1)) = \frac{1}{L}\sum_{l=1}^{L} B_{12}^N(\mathbf{x}(l)) \quad .$$

Notice that the right hand side of this expression is a U-statistics, as pointed out by Dmochowsky (1994, Theorem 2), for exchangeable observations. Then

$$\frac{1}{L}\sum_{l=1}^{L} B_{12}^N(\mathbf{x}(l)) \to E_\theta^M\left[\frac{m_1^N(\mathbf{x}(l))}{m_2^N(\mathbf{x}(l))}\right] = \int \frac{m_1^N(\mathbf{x}(l))}{m_2^N(\mathbf{x}(l))} f(\mathbf{x}(l))|\theta)d\mathbf{x}(l) \quad ,$$

where $f(\mathbf{x}(l))|\theta)$ is the true sampling density. For π_1 and π_2 to produce a Bayes factor that tends to the arithmetic IBF with n, and assuming that either M_1 or M_2 is the sampling model, the priors must obey the following equations: under M_1,

$$\frac{\pi_2(\psi_2(\theta_1))\pi_1^N(\theta_1)}{\pi_2^N(\psi_2(\theta_1))\pi_1(\theta_1)} = E_{\theta_1}^{M_1}[B_{12}^N(\mathbf{x}(l))] \tag{4}$$

and, under M_2,

$$\frac{\pi_2(\theta_2)\pi_1^N(\psi_1(\theta_2))}{\pi_2^N(\theta_2)\pi_1(\psi_1(\theta_2))} = E_{\theta_2}^{M_2}[B_{12}^N(\mathbf{x}(l))]$$

where $\psi_i(\theta_j)$ denotes the limit of the MLE of model i when model j is the true one.

To analyse the nested model situation we first consider some notation: we shall say that M_1 and M_2 are nested if $\theta_2 = (\xi, \eta)$ and $f_1(x \mid \theta_1, \eta_0) = f_2(x \mid \xi = \theta_1, \eta = \eta_0)$ where η_0 is a specified value of n. Observe that $\phi_2(\theta_1) = (\theta_1, \eta_0)$, thus (4) is redundant and, as noted in Dmochowski (1994), it is seen that the general solutions to the instrinsic equation are

$$\left\{\begin{array}{lcl} \pi_1^I(\theta_1) & = & \pi_1^N(\theta_1)u(\theta_1) \\ \pi_2^I(\xi,\eta) & = & \pi_2^N(\xi,\eta)E_{(\xi,\eta)}^{M_2}[B_{12}^N(\mathbf{x}(l))]u(\psi_1(\xi,\eta)) \end{array}\right. \tag{5}$$

where an I superscript denotes the intrinsic prior and u is an arbitrary nonnegative continuous function.

Equation (5) shows that a particular intrinsic prior depends on a particular choice of a function u, thus the class of intrinsic priors can be set as the class of solutions of (5) when u ranges in the class of continuous functions.

In Berger and Pericchi (1995) the idea of intrinsic priors was first introduced by studying the simplest solution to (5) which is obtained when $u \equiv 1$. In that paper the authors prove that if $\pi_1^N(\theta_1)$ is proper then so is $\pi_2^I(\xi, \eta)$ and it is shown in several examples that even if $\pi_1^N(\theta_1)$ is the improper reference prior, $\pi_2^I(\eta|\xi)$ is proper.

2.1 Properties of the class of intrinsic priors. The following remarks illustrate some properties of the intrinsic priors in the nested model situation:

REMARK 1 Once a specific u is set, it fixes both $\pi_1^I(\theta_1)$ and $\pi_2^I(\xi,\eta)$. Thus (5) establishes automatically a relashionship between the two priors. Notice the prominent röle played by $E^{M_2}_{(\xi,\eta)}[B_{12}(\mathbf{x}(l))]$ and $\psi_1(\xi,\eta))$ in such relationship.

REMARK 2 Given the large choice of possible functions u, the class of intrinsic priors is truly big. This anticipates a general robustness result: the arithmetic IBF can be considered robust, with respect to large classes of priors on θ_1, provided that the corresponding $\pi_2(\xi,\eta)$ are assigned according to (5).

REMARK 3 Given two solutions of (5), say $(\pi_1^{I,0},\pi_2^{I,0})$ and $(\pi_1^{I,1},\pi_2^{I,1})$, any convex combination $(\lambda\pi_1^{I,0}+(1-\lambda)\pi_1^{I,1},\lambda\pi_2^{I,0}+(1-\lambda)\pi_2^{I,1})$, $\lambda\in[0,1]$, is also a solution. This result holds also for separate models and establishes that the class of intrinsic priors is convex.

Notice that when a simple null hypothesis with no free parameter is considered, that is $M_1:\eta=\eta_0$ and $M_2:\eta\neq\eta_0$ the only possible prior under M_1 is that which concentrates all its mass in the null hypothesis, thus there is only one solution to the intrinsic equations and (it can be proved that) it is always proper.

2.2 Intrinsic priors and expected IBF. The *expected* IBF is obtained from (2) when the average is substituted by the expected value of $B_{12}^N(x(\ell))$ under M_2, yielding

$$B_{21}^{EAI}=B_{21}^N E^{M_2}_{(\hat\xi,\hat\eta)}[B_{12}(\mathbf{x}(l))] \tag{6}$$

To explore the relationship between this and (5) write

$$B_{21}=\frac{\displaystyle\int\int f_2(\mathbf{x}|\xi,\eta)\frac{\pi_2^I(\xi,\eta)}{\pi_2^N(\xi,\eta)}\pi_2^N(\xi,\eta)d\xi d\eta}{\displaystyle\int f_1(\mathbf{x}|\theta_1)\frac{\pi_1^I(\theta_1)}{\pi_1^N(\theta_1)}\pi_1^N(\theta_1)d\theta_1} \tag{7}$$

this yields the approximation

$$B_{21}\approx B_{21}^N E^{M_2}_{(\hat\xi,\hat\eta)}[B_{12}(\mathbf{x}(l))]\frac{u(\psi_1(\hat\xi,\hat\eta))}{u(\hat\theta_1)} \tag{8}$$

where the hats denote the MLE under the corresponding model.

The approximation in (8) may seem odd, since it appears to be different to (6) whenever u is not a constant. However both approximations coincide, as the following lemma shows.

LEMMA 1 *Assume that the likelihood f_1 is unimodal and that M_1 is nested in M_2. Denote by Θ_1 the range of θ_1 and by Θ_2 the range of (ξ, η). Then*

$$\psi_1(\hat{\xi}, \hat{\eta}) = \hat{\theta}_1 \quad .$$

Proof: First notice that

$$\{\theta_1 = \psi_1(\xi, \eta) : (\xi, \eta) \in \Theta_2\} \subset \Theta_1 \tag{9}$$

by definition of ψ_1. On the other hand $\psi_1(\xi = \theta_1, \eta_0) = \theta_1$, since M_1 is nested in M_2, therefore the inclusion in (9) is an equality, and thus ψ_1 is onto. On the other hand, since f_1 is unimodal, then $\hat{\psi_1}(\xi, \eta) = \hat{\theta}_1$. Finally, due to the invariance of the MLE, $\hat{\psi_1}(\xi, \eta) = \psi_1(\hat{\xi}, \hat{\eta})$, which completes the proof.

2.3 Robustness for fixed sample size. In 2.2 we have seen that the Bayes factor corresponding to priors within the class of intrinsic priors, tends to B_{21}^{EAI} as $n \to \infty$. The following result is related to the fixed sample case.

LEMMA 2 *Suppose that f_1 and f_2 are such that $f_2(x \mid \xi, \eta) = f_1(x \mid \xi, \eta_0) g_2(x \mid \eta)$ for some g_2, $\psi_1(\xi, \eta) = \theta_1, E_{(\xi,\eta)}^{M_2}[B_{12}(x(l))] = h(\eta)$, for some h, and $\pi_2^N(\xi, \eta) = \pi_i^N(\xi)\pi_{22}^N(\eta)$, then any solution of (5) will produce the same Bayes factor, independently of the choice of u.*

PROOF: The conditions imply that $\pi_2^I(\xi, \eta) = \pi_1^I(\xi)\pi_{22}^I(\eta)$, so Fubini's theorem can be used to see that $B_{21} = \int g_2(x \mid \eta)\pi_{22}^I(\eta) d\eta$.

The following model illustrates the result. Consider $y_1 \dots y_{2n}$ an i.i.d. normal sample of size $2n$ such that $y_i \sim N(\theta + \delta/2, 1)$ for $i = 1, \dots, n$ and $y_i \sim N(\theta - \delta/2, 1)$ for $i = n+1, \dots, 2n$. Let M_1 be the model that corresponds to $\delta = 0$ and $\pi_1^N(\theta) \equiv 1$ and M_2 the model that corresponds to $\delta \neq 0$ and $\pi_2^N(\theta, \delta) \equiv 1$. Calculations for the Bayes factor yield

$$B_{21}^N = \frac{2\sqrt{\pi}}{\sqrt{n}} \exp\left(\frac{n}{4}(\overline{y}_2 - \overline{y}_1)^2\right)$$

where $\overline{y}_i$ is the average of the ith group. The minimun training sample $\mathbf{y}(l)$ is a subsample of size 2 containing one observation per group and

$$E_{(\theta,\delta)}^{M_2}[B_{12}^N(\mathbf{y}(l))] = \frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{\delta^2}{8}\right) \quad .$$

On the other hand the MLE of θ under M_1 is $(\sum_{i=1}^{n} y_i+\sum_{i=n+1}^{2n} y_i)/(2n)$ and, when M_2 is the right model, $(\sum_{i=1}^{n} y_i)/n \to \theta+\delta/2$ and $(\sum_{i=n+1}^{2n} y_i)/n \to \theta-\delta/2$, thus $\psi_1(\theta,\delta)=\theta$, implying that the intrinsic priors are $\pi_1^I(\theta)=u(\theta)$ and $\pi_2^I(\theta,\delta)=u(\theta)N(\delta|0,4)$. For this choice of priors the Bayes factor, B_{21}, is a ratio where the denominator is

$$D=\int_{-\infty}^{\infty}\exp\left(-\sum_{i=1}^{2n}(y_i-\theta)^2/2\right)u(\theta)d\theta$$

and the numerator is

$$D\int_{-\infty}^{\infty}\exp\left(-\frac{n}{4}((\delta+(\overline{y}_2-\overline{y}_1))^2+(\overline{y}_2-\overline{y}_1)^2)\right)N(\delta|0,4)d\delta \quad ,$$

so that B_{21} does not depend on the choice of u for any sample size.

2.4 Non-parametric classes. An analysis of the robustness of B_{21} for a finite sample size can be achieved by considering classes of priors within the class of solutions of (5). One possibility is to consider band classes. Suppose that the change of variables $(\xi,\eta)\mapsto(t,s)$ with $t=\psi_1(\xi,\eta)$ is one to one, then (8) yields

$$B_{21}=\frac{\int u(t)I(t)dt}{\int u(\theta_1)k_1(\theta_1)d\theta_1}$$

for appropriate $I(t)$ and k_1. B_{21} is thus expressed as a ratio of linear functionals of u and, if θ_1 is a real parameter, it is possible to obtain the sup and inf of B_{21} within a class, say $\mathcal{U}=\{u(t):a(t)\le u(t)\le b(t)\}$ for some functions $a(t)$ and $b(t)$. Note however that such a class allows wildly discontinuous functions u, an undesirable feature that may destroy the identity between the right hand sides of (8) and (6).

Given the convexity of the class of intrinsic priors, it is very natural to explore its robustness in terms of convex combinations of priors. Given two solutions of (5), say $(\pi_1^{I,0},\pi_2^{I,0})$ and $(\pi_1^{I,1},\pi_2^{I,1})$, that correspond respectively to two functions u_0 and u_1, consider the class

$$\mathcal{C}=\{(\pi_1,\pi_2)=(\lambda\pi_1^{I,0}+(1-\lambda)\pi_1^{I,1},\lambda\pi_2^{I,0}+(1-\lambda)\pi_2^{I,1})\ ,\ \lambda\in[0,1]\} \quad .$$

It is easily seen that the maximum and the minimum of B_{21} are attained at the extremes of the class.

As an illustrative example consider the normal location-scale problem, that is a set of n independent observations having density $N(0,\sigma_1^2)$ and prior $\pi_1(\sigma_1)=\sigma_1^{-1}$ under M_1 and density $N(\theta,\sigma_2)$ with prior $\pi_2(\theta,\sigma_2^2)=\sigma_2^{-2}$ under M_2. It is proved in Berger and Pericchi (1995) that $\psi_1(\theta,\sigma)=\theta^2+\sigma^2$ and

$$E_{(\theta,\sigma)}^{M_2}[B_{12}(\mathbf{x}(l))]=\frac{1-\exp(-\theta^2/\sigma^2)}{2\sqrt{\pi}(\theta^2/\sigma^2)} \quad .$$

Consider $u_0(\sigma^2) \equiv 1$ and $u_1(\sigma^2) = \exp(-1/(\beta\sigma^2))/(\beta^\alpha\sigma^{2\alpha}\Gamma(\alpha))$ so that $\pi_1^{I,0} = \pi_1^N$ and $\pi_1^{I,1}$ is an inverted gamma. This choice of u_1 guarantees that both $\pi_1^{I,1}$ and $\pi_2^{I,1}$ are proper densities. Table 1 shows the values of B_{21}^0, B_{21}^1 and B_{12}^{EIA} for imaginary samples with $\overline{x} = 0.1$ and $\overline{x} = 0.5$ supposing $s^2 = \beta = \alpha = 1$. The results show a very robust behaviour of the Bayes factor for this class.

3. The Expected IBF is more robust than proper Bayes factors. Large classes of priors for the parameters under test, produce lack of robustness. Consider an i.i.d. sample of n normally distributed obervations and let $y_i \sim N(0,1)$ for $i = 1 \ldots n$ under M_1 and $y_i \sim N(\theta,1)$ under M_2, $\theta \neq 0$ and $\pi_2(\theta) = N(\theta|0,\tau^2)$. Clearly in this case no non-informative prior is used and so no correction term for the Bayes factor should be needed. It has been noticed in Berger (1985, p. 151) that the resulting Bayes factor for comparing M_1 with M_2 lacks robustness with respect to τ^2, an annoying feature, since the value of τ^2 is typically difficult to elicit. In fact

$$B_{21} = \exp\left(\frac{\overline{y}^2 n^2 \tau^2}{2(n\tau^2+1)}\right)\frac{1}{\sqrt{1+n\tau^2}}$$

and thus $B_{21} \to 0$ when $\tau^2 \to \infty$. Note that this is the case for any fixed n, so B_{21} is not robust even asymptotically. However, consider and imaginary observation $y \sim N(\theta,1)$, then, still using $\pi_2(\theta)$ as above.

$$E_\theta^{M_2}[B_{12}(y)] = \frac{1+\tau^2}{\sqrt{2\tau^2+1}}\exp\left(-\frac{\theta^2\tau^2}{2(2\tau^2+1)}\right)$$

and this implies that the expected IBF is

$$B_{21}^{EAI} = \frac{1+\tau^2}{\sqrt{(2\tau^2+1)(1+n\tau^2)}}\exp\left(\frac{\overline{y}^2\tau^2(n^2\tau^2(2n-1)+n^2-1)}{2(n\tau^2+1)(2\tau^2+1)}\right) \quad ;$$

TABLE 1: *Upper and lower bound of B_{21} for the normal location-scale problem*

		$n = 10$	$n = 20$	$n = 35$	$n = 50$
$\overline{x} = 0$	B_{21}^0	0.2134	0.1542	0.1178	0.0990
	B_{21}^1	0.2134	0.1543	0.1179	0.0990
	B_{21}^{EAI}	0.2236	0.1581	0.1195	0.1000
$\overline{x} = 0.5$	B_{21}^0	0.6468	1.416	5.747	25.70
	B_{21}^1	0.6467	1.417	5.749	25.71
	B_{21}^{EAI}	0.6650	1.438	5.798	25.86

since $\lim_{\tau^2\to\infty} B_{21}^{EAI} = \exp(\overline{y}^2 n/2)/\sqrt{2n} > 0$, τ^2 has a bounded influence over B_{21}^{EAI}. Figure 1 illustrates the behaviour of B_{21}^{EAI} and B_{21} for values of $\tau^2 \geq 1$. The corresponding intrinsic prior

$$\pi^I(\theta) = \frac{1}{\sqrt{2\pi}} \frac{1+\tau^2}{\sqrt{(2\tau^2+1)\tau^2}} \exp\left(-\frac{\theta^2(\tau^4+2\tau^2+1)}{2\tau^2(2\tau^2+1)}\right)$$

is a $N(\theta|0, \tau^2(2\tau^2+1)/(\tau^4+2\tau^2+1))$, so that the original variance τ^2 is reduced by a factor of $(2\tau^2+1)/(\tau^4+2\tau^2+1)$ and this achieves the desired robustness. Using k observations the resulting reduction is $(2k\tau^2+1)/(k^2\tau^4+2k\tau^2+1))$, since robustness is already achieved for $k=1$, the later seems to be the optimal choice. Note that, when $\tau^2 \to \infty$, the intrinsic prior tends to a $N(\theta|0,2)$ and this corresponds to the intrinsic prior that is obtained when a uniform prior is assumed and a minimal training sample is considered.

This robustifying effect of the IBF strategy is quite a general feature in problems where, given a class of priors $\pi_\tau(\theta)$ $\tau \in \mathcal{T}$ and a reference prior $\pi_R(\theta)$, $\pi_\tau(\theta|\mathbf{x}(l)) \xrightarrow{\mathcal{T}} \pi_R(\theta|\mathbf{x}(l))$. The reason being that for a reference prior the expected IBF is neither zero nor infinity, but depends on the data at hand.

4. Change in prior measures. Consider the normal location-scale problem as in the previous section letting the more general choice of priors $\pi_1^N(\sigma_1) = \sigma_1^{-(1+r)}$ under M_1 and $\pi_2^N(\theta,\sigma_2) = \sigma_2^{-(1+q)}$ under M_2. This problem can be considered as a particular case of the problem of comparing nested linear models analysed in Berger and Pericchi (1994), the analysis below is indeed valid for the linear models case, but it will be presented in terms of the normal location-scale problem for the sake of clarity.

It is then clear that the minimum training sample consists of two different observations and that, letting $\overline{x} = \sum x_i/n$ and $s^2 = \sum(x_i-\overline{x})^2/(n-1)$,

$$B_{21}^N = \frac{m_2^N(x)}{m_1^N(x)} = \sqrt{\frac{2\pi}{n}}\,2^{(q-r-1)/2}\frac{\Gamma\left(\frac{n+q-1}{2}\right)\left(\sum x_i^2\right)^{(n+r)/2}}{\Gamma\left(\frac{n+r}{2}\right)\left((n-1)s^2\right)^{(n+q-1)/2}}.$$

In the appendix it is shown that,

$$E^{M_2}_{(\theta,\sigma_2^2)}[B_{12}(l)] = \frac{e^{-\lambda/2}}{\sigma_2^{1-q+r}}\frac{\Gamma\left(\frac{q+1-r}{2}\right)\Gamma\left(\frac{q+2}{2}\right)\Gamma\left(\frac{r+2}{2}\right)}{\Gamma\left(\frac{q+1}{2}\right)\Gamma\left(\frac{q+3}{2}\right)\pi\sqrt{2}} M\left(\frac{q+1-r}{2}, \frac{q+3}{2}, \frac{\lambda}{2}\right)$$

(10)

where $\lambda = 2\theta^2/\sigma_2^2$ and $M(a,b,c)$ denotes Kummer's function (see, for example, Abramowitz and Stegun, 1965, Chapter 13). Thus the intrinsic equations are

$$\begin{cases} \pi_1^I(\sigma_1) &= \sigma_1^{-(r+1)} u(\sigma_1^2) \\ \pi_2^I(\theta,\sigma_2) &= \sigma_2^{-(q+1)} \frac{K_{qr}\sigma_2^{q-r-1}}{\pi\sqrt{2}} e^{-\lambda/2} M\left(\frac{q+1-r}{2}, \frac{q+1}{2}, \frac{\theta^2}{\sigma_2^2}\right) u(\theta^2+\sigma_2^2) \end{cases}$$

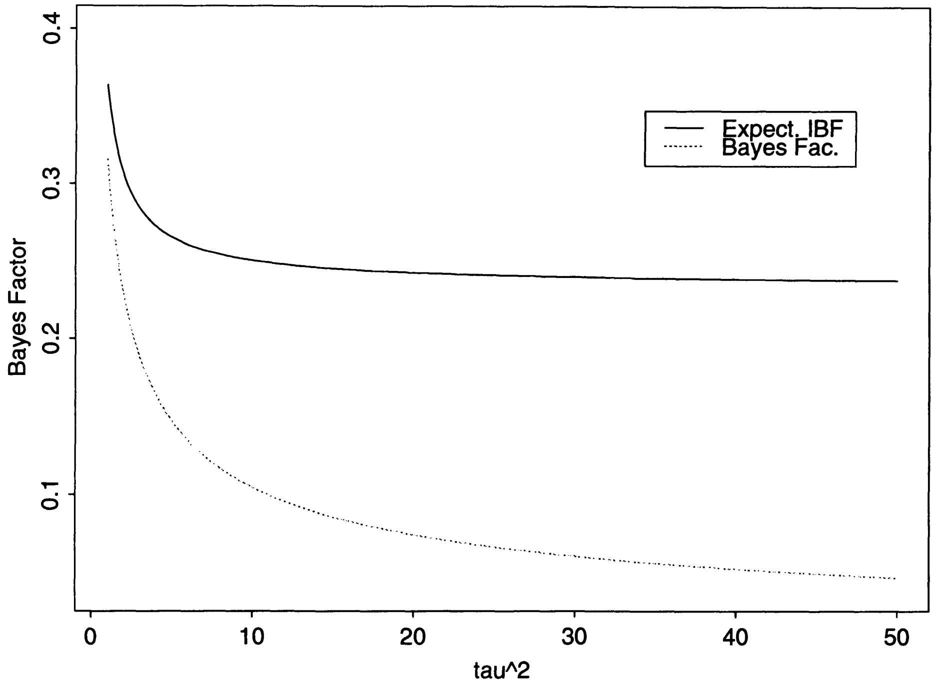

Figure 1: *Bayes factor and expected IBF as a function of* τ^2 ($\bar{y} = .1 \ , n = 10$)

where

$$K_{qr} = \frac{\Gamma((r+2)/2)\Gamma(q+2/2)\Gamma(q+1-r)/2)}{\Gamma(q+1)/2)\Gamma(q+3)/2)}$$

and it is clear that when $u \equiv 1$, π_1^I is an improper density. Thus the question arises, is

$$\text{(11)} \qquad \pi_2^I(\theta|\sigma_2^2) = \frac{K_{qr}\sigma_2^{-1}}{\pi\sqrt{2}} e^{-\theta^2/\sigma_2^2} M\left(\frac{q+1-r}{2}, \frac{q+3}{2}, \frac{\theta^2}{\sigma_2^2}\right)$$

a proper density for $u \equiv 1$? This is an important question, since at least the prior for the parameter under test should be well calibrated for $\pi_1^I(\sigma) = \pi_2^I(\sigma)$.

The answer to this question is, in general, no, since, for $q > -1$,

$$\int_{-\infty}^{\infty} \pi_2^I(\theta|\sigma_2^2)d\theta = \frac{\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{q-r+1}{2}\right)}{\Gamma\left(\frac{q+1}{2}\right)\sqrt{\pi}}$$

(see the appendix) and this expression is equal to one only if either $r = q$ or $r = 0$. This suggests that the intrinsic prior is well calibrated if either both models have essentially the same prior for the scale parameter or the reference prior is used for the null model. A remarkable feature of the former identity is that it holds for linear models of general dimensions, as can be seen in the appendix, and the constant does not depend on the dimension of the regressors.

It is possible to gain knowledge about the behaviour of $\pi_2(\theta|\sigma_2^2)$ by analysing the asymptotic expansion of M. It is then seen that r governs the tails of the intrinsic prior, in fact $\pi_2(\theta|\sigma_2^2)$ behaves as a Student-t with $r+1$ degrees of freedom for large θ.

To assess the robustness of the IBF with respect to the choice of q and r, consider the expected IBF, that is B_{21}^N times formula (10) evaluated at $\lambda = 2n\overline{x}^2/((n-1)s^2)$. Then, using the fact that the Kummer's function can be written as the moment generating function of a beta distribution (see, Abramowitz and Stegun, 1965, 13.2.1) the following bound can be obtained

$$\exp\left(-\frac{\overline{x}^2 n}{\sum(x_i-\overline{x})^2}\right) C \leq B_{21}^N E_{(\hat{\theta},\hat{\sigma}_2^2)}^{M_2}[B_{12}(x(l))] \leq C$$

where

$$C = \left(\frac{2}{n}\right)^{(q-r)/2} \sqrt{2\pi}\left(\frac{\sum x_i^2}{(n-1)s^2}\right)^{(n+r)/2} \frac{\Gamma\left(\frac{n+q-1}{1}\right)}{\Gamma\left(\frac{n+r}{2}\right)} K_{qr} \quad .$$

Using Stirling's formula to approximate the Gamma function it is seen that, when $r = 0$ and $q \to \infty$ then $C \to \infty$ and thus the expected IBF tends to infinity, for any value of the data. When $r = q$ the situation is different since the behaviour of C for large q depends on $t^2/(n-1) = n\overline{x}^2/((n-1)s^2)$. In this case it is seen that $C \to \infty$ except when $t^2/(n-1)$ is very small and so the data strongly favors the null hypothesis; in such cases $C \to 0$ as $q \to \infty$.

In the light of the preceeding results, extreme prior measures (large q or r) are clearly undesirable. Given that r governs the tails of the intrinsic prior, the choice of $r = 0$ and q small seems best.

4.1 Are intrinsic priors well calibrated? In the tradition of Jeffreys' conventional priors (Jeffreys (1961, Ch. 5)), a reasonable proper prior is assigned for the parameter under test, conditional on the other parameters, and a reference improper prior is employed for the 'common' parameters. An argument for this is that, since the parameters are 'common' to both models, their proportionality constants cancel out in the Bayes factor. This argument is suspect, for 'common' parameters may have different meaning in different models.

We approach the problem following the IBF strategy of calibrating the priors by minimal training samples, as follows. Let $\pi(\eta|\xi)$ be a proper density and $\pi(\xi) = \pi(\theta_1)$ improper. Consider a minimal training sample $\mathbf{x}(l)$, so that both $\pi(\xi, \eta|\mathbf{x}(l))$ and $\pi(\theta_1|\mathbf{x}(l))$ are proper. The conditional proper prior approach will be regarded as well calibrated whenever the correction factor generated by $\mathbf{x}(l)$ cancels out in the Bayes factor.

Let us illustrate the approach with the normal location-scale problem. We have seen that starting with $\pi_1^N(\sigma_1) = \sigma_1^{-(1+r)}$ and $\pi_2 = \sigma_2^{-(1+q)}$, for $r = 0$ and $q > -1$ or $r = q \geq 0$, we obtain $\pi_2^I(\theta|\sigma_2) = E^{M_2}[B_{12}(\mathbf{x}(l))]\sigma_2^{q-r}$ as in (11). Let now the priors be $\pi_1(\sigma_1) = \sigma_1^{-(r+1)}$ and $\pi_2(\sigma_2) = \sigma_2^{-(r+1)}$ (note that $\pi_1(\sigma_1) = \pi_1^N(\sigma_1)$ and $\pi_2(\sigma_2) = \pi_2^N(\sigma_2)\sigma_2^{q-r}$). Since $\pi_2^I(\theta|\sigma_2)$ is a proper density the minimum training sample is of size one and

$$B_{12}^N(x(l)) = \frac{\displaystyle\int_0^\infty (2\pi\sigma_1^2)^{1/2} \exp(-x(l)^2/(2\sigma_1^2))\sigma_1^{-(1+r)} d\sigma_1}{\displaystyle\int_0^\infty m_2(x(l)|\sigma_2)\sigma_2^{-(1+r)} d\sigma_2}$$

where

$$m_2(x(l)|\sigma_2) = \frac{K_{qr} e^{-x^2/(3\sigma_2^2)}}{\pi\sqrt{2}\sqrt{2\pi}\sigma_2^2}$$
$$\int_{-\infty}^\infty \exp\left(-\frac{3(\theta - x(l)/3)^2}{2\sigma_2^2}\right) M\left(\frac{q+1-r}{2}, \frac{q+3}{2}, \frac{\theta^2}{\sigma_2^2}\right) d\theta \quad .$$

Using the definition of M, $m_2(x(l)$ can be written as

$$K_{qr} \frac{e^{-x(l)^2/(3\sigma_2^2)}\sigma_2^{-1}}{\sqrt{2\pi}\sqrt{3}} \frac{\Gamma(q+3/2)}{\Gamma(q+1-r)/2)} \sum_{j=1}^\infty \frac{\Gamma(q+1-r+2j)/2)}{\Gamma(q+3+2j)/2)} \frac{1}{j!} E\left[\frac{\theta}{\sigma_2}\right]^{2j} \quad ,$$

where the expectation is taken with respect to a normal with mean $x(l)/3$ and variance $\sigma_2^2/3$. $E(\theta/\sigma_2)^{2j}$ is a function of $x(l)^2/\sigma_2^2$ and so $m_2(x(l)|\sigma_2)$ can be written as $\sigma_2^{-1}H_{qr}(x(l)^2/\sigma_2^2)/\sqrt{2\pi}$ where

$$\begin{aligned} H_{qr}(x(l)^2/\sigma_2^2) \quad &= \quad K_{qr} \frac{e^{-x(l)^2/(3\sigma_2^2)}}{\sqrt{3}} \frac{\Gamma((q+3)/2)}{\Gamma((q+1-r)/2)} \\ &\quad \sum_{j=1}^\infty \frac{\Gamma(q+1-r+2j)/2)}{\Gamma(q+3+2j)/2)} \frac{1}{j!} E\left[\frac{\theta}{\sigma_2}\right]^{2j} \quad , \end{aligned}$$

thus, since $\int m_2(x(l)|\sigma_2)dx(l) = 1$, (recall that $\pi_2^I(\theta|\sigma_2)$ is proper) $m_2(x(l)|\sigma_2)$ is a symmetric scale likelihood for $x(l)$.

The former discussion leads to

$$B_{12}^N(x(l)) = \frac{\displaystyle\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{x(l)^2}{2\sigma_1^2}\right) \sigma_1^{-(1+r)} d\sigma_1}{\displaystyle\int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_2} H_{qr}\left(\frac{x(l)^2}{\sigma_2^2}\right) \sigma_2^{-(1+r)} d\sigma_2},$$

and, for the correction factor of the IBF to be one, each of the $B_{12}^N(x(l))$ has to be one, that is

$$\int_0^\infty \frac{1}{\sigma_2} H_{qr}\left(\frac{x(l)^2}{\sigma_2^2}\right) \sigma_2^{-(1+r)} d\sigma_2 = \frac{\Gamma((r+1)2)}{2} \left(\frac{x(l)^2}{2}\right)^{-(r+1)/2}.$$

The previous condition is naturally satisfied using lemma 2 when $r = 0$ for any q. Thus the use of $\pi_1(\sigma) = 1/\sigma_1$ and $\pi_2(\theta, \sigma_2) = 1/\sigma_2 \pi_2^I(\theta|\sigma_2)$ guarantees that the comparison of M_1 with M_2 is well calibrated. It is still being studied if for $r = q > 0$ the above identity holds.

5. Open problems and conclusions. The issues discussed in this paper have led us to the following conclusions: 1) The IBF method is a way to automatically elicit a class that is fairly robust with respect to the problem of comparing two nested models. The class is convex and the intrinsic equations establish the conditions that the prios must satisfy so that asymptotic robustness holds. When there are free 'common' parameters the class appears to be rich. 2) Bayesian model comparisons are inherently non-robust with respect to wide classes of priors of the parameters under test. However the expected IBF is considerably more robust than proper Bayes factors, for large classes of priors. 3) Various robustness considerations point out that, for the linear model, using the reference prior for the null model is an optimal choice. Is this the case in more general settings?

Important future tasks are the study of robust comparisons for separate models and for mixtures of several models.

Acknowledgements. The authors are grateful to Prof. Jim Berger and to Prof. Alain Etcheberry.

Appendix

Consider the models $M_j : \mathbf{Y} = \mathbf{X}_j\beta_j + \varepsilon_j$ where $\varepsilon_j \sim N_n(0, \sigma_j^2\mathbf{I}_n)$. $\beta_j \in \Re^{k_j}$ and σ_j are unknown and $\mathbf{X}_j$ is the $(n \times k_j)$ design matrix $(k_j < n)$. $\hat{\beta}_j = (\mathbf{X}_j^t\mathbf{X}_j)^{-1}\mathbf{X}_j^t\mathbf{Y}$ is the least squares estimator of β_j and $R_j = |\mathbf{Y} - \mathbf{X}_j\hat{\beta}_j|^2$ is the residual sum of squares. Consider default priors $\pi_j^N(\beta_j, \sigma_j) =$

$\sigma_j^{-(1+q_j)}$, $q_j > -1$. The following results are a generalisation of those obtained in Berger and Pericchi (1994) for specific values of q_j.

To compare M_1 with M_2 ($k_2 > k_1$), the correcting factor is

$$\frac{C}{L}\sum_{l=1}^{L}\frac{|\mathbf{X}_2^t(l)\mathbf{X}_2(l)|^{1/2}R_2(l)^{(q+1)/2}}{|\mathbf{X}_1^t(l)\mathbf{X}_1(l)|^{1/2}R_1(l)^{(r+p+1)/2}} \tag{12}$$

where $q_2 = q$ and $q_1 = r$, $C = \pi^{-p/2}\Gamma((r+p+1)/2)/(2^{(q-r)/2}\Gamma((q+1)/2))$ and $p = k_2 - k_1$. Write $\beta_2^t = (\beta_0, \beta^*)^t$ where $\beta_0 \in \Re^{k_1}$ corresponds to β_1 under M_1. To obtain $\pi_2(\beta^*|\beta_0, \sigma_2)$ and the expected IBF, the expectation of (12) is needed, to this end note that

$$E^{M_2}\left[\frac{R_2(l)^{(q+1)/2}}{R_1(l)^{(r+p+1)/2}}\right] = E^{M_2}\left[\frac{W^{(q+1)/2}}{(W+V)^{(r+p+1)/2}}\right]\sigma_2^{(q-p-r)}$$

where W and V are independent and $W \sim \chi_1^2$ and $V \sim \chi_p^2(\lambda_l(\beta^*))$, a non-central chi square distribution with non-centrality parameter

$$\lambda_l(\beta^*) = \sigma_2^{-2}\beta_2^t\mathbf{X}_2^t(l)\left(\mathbf{I} - \mathbf{X}_1(l)[\mathbf{X}_1^t(l)\mathbf{X}_1(l)]^{-1}\mathbf{X}_1^t(l)\right)\mathbf{X}_2(l)\beta_2$$

as defined in Berger and Pericchi (1994).

Note that

$$E^{M_2}\left[\frac{W^{(q+1)/2}}{(W+V)^{(r+p+1)/2}}\right] = \sum_{k=0}^{\infty}\left(\frac{\lambda_l(\beta^*)}{2}\right)^k\frac{e^{-\lambda_l(\beta^*)/2}}{k!}T_k$$

where

$$T_k = E^{M_2}\left[\frac{W^{(q+1)/2}}{(W+\chi^2_{p+2k})^{(r+p+1)/2}}\right] = \frac{2^{(q-p-r)/2}}{\sqrt{\pi}}\frac{\Gamma\left(\frac{q+2}{2}\right)\Gamma\left(\frac{q+2k+1-r}{2}\right)}{\Gamma\left(\frac{2k+p+q+2}{2}\right)},$$

then the expectation of (12) is

$$\frac{C\sigma_2^{q-p-r}\Gamma\left(\frac{q+2}{2}\right)}{2^{-(q-p-r)/2}L\sqrt{\pi}}\sum_{l=1}^{L}\frac{|\mathbf{X}_2^t(l)\mathbf{X}_2(l)|^{1/2}}{|\mathbf{X}_1^t(l)\mathbf{X}_1(l)|^{1/2}}e^{-\frac{\lambda_l(\beta^*)}{2}}\sum_{k=0}^{\infty}\left(\frac{\lambda_l(\beta^*)}{2}\right)^k\frac{1}{k!}\frac{\Gamma\left(\frac{q+2k+1-r}{2}\right)}{\Gamma\left(\frac{2k+p+q+2}{2}\right)}$$

and, using the definition a Kummer's function M, this last expression is

$$\frac{C\sigma_2^{q-p-r}\Gamma\left(\frac{q+2}{2}\right)\Gamma\left(\frac{q+1-r}{2}\right)}{2^{-(q-p-r)/2}L\sqrt{\pi}\Gamma\left(\frac{q+p+2}{2}\right)}$$

$$\sum_{l=1}^{L}\frac{|\mathbf{X}_2^t(l)\mathbf{X}_2(l)|^{1/2}}{|\mathbf{X}_1^t(l)\mathbf{X}_1(l)|^{1/2}}e^{-\lambda_l(\beta^*)/2}M\left(\frac{q+1-r}{2},\frac{q+p+2}{2},\frac{\lambda_l(\beta^*)}{2}\right). \tag{13}$$

Now, $\pi_2(\beta^*|\beta_0,\sigma_2)$ is equal to (13) multiplied by σ_2^{r-q} and an obvious choice of $g_l(\beta^*)$ gives $\pi_2(\beta^*|\beta_0,\sigma_2) = 1/L\sum_{l=1}^L g_l(\beta^*)$. To calculate $\int g_l(\beta^*)d\beta^*$ consider the change of variables $\beta^* \mapsto \lambda_l(\beta^*)$. This has a Jacobian

$$\frac{|\mathbf{X}_1^t(l)\mathbf{X}_1(l)|^{1/2}}{|\mathbf{X}_2^t(l)\mathbf{X}_2(l)|^{1/2}}\frac{(\pi\sigma_2^2)^{p/2}}{\Gamma(p/2)}\lambda_l(\beta^*)^{(p-2)/2}$$

and thus

$$\begin{aligned}
&\int g_l(\beta^*)d\beta^* \\
&= \frac{\sigma_2^{q-r}\Gamma\left(\frac{r+p+1}{2}\right)\Gamma\left(\frac{q+2}{2}\right)}{2^{p/2}\sqrt{\pi}\Gamma\left(\frac{q+1}{2}\right)\Gamma\left(\frac{p}{2}\right)}\sum_{k=0}^{\infty}\frac{\Gamma\left(\frac{q+1-r+2k}{2}\right)2^{(k+p/2)}}{\Gamma\left(\frac{q+p+2+2k}{2}\right)k!2^k}\Gamma\left(\frac{2k+p}{2}\right) \\
&= \frac{\sigma_2^{q-r}\Gamma\left(\frac{r+p+1}{2}\right)\Gamma\left(\frac{q+2}{2}\right)\Gamma\left(\frac{q+1-r}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{q+1}{2}\right)\Gamma\left(\frac{q+p+2}{2}\right)}F\left(\frac{q+1-r}{2},\frac{p}{2},\frac{q+p+2}{2},1\right) \\
&= \frac{\sigma_2^{q-r}\Gamma\left(\frac{q+1-r}{2}\right)\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{q+1}{2}\right)},
\end{aligned}$$

where F is the hypergeometric function, and 15.1.20 in Abramowitz and Stegun (1965) has been used.

REFERENCES

Albert, J., and Chib, S. (1994). Bayesian model diagnostics in hierarchical models, Preprint.

Abramowitz, M. and Stegun, I. (1965) (eds) *Handbook of Mathematical Functions.* New York: Dover.

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

Berger, J.O. and Pericchi, L.R. (1994). The intrinsic Bayes factor for linear models. Technical Report 94-10C. Purdue University, Department of Statistics.

Berger, J.O. and Pericchi, L.R. (1995). The intrinsic Bayes factor for model selection and prediction. *J. of the Amer. Statist. Assoc.*: In press.

De Santis, F. and Spezzaferri, F. (1995). Comparing hierarchical models using Bayes factor and fractional Bayes factor.: A robust analysis. Unpublished.

Dmochowski, J. (1994). Intrinsic priors via Kullback-Liebler geometry.

Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. (eds) *Bayesian Statistics 5*, 543-549.

Jeffreys, H. (1961). *Theory of Probability*, third edn. Oxford: Oxford University Press.

O'Hagan, A. (1995) Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society*, B, 57, 99-138: (with discussion).

Bruno Sansó
Luis R. Pericchi
CESMa
Universidad Simón Bolivar
Apartado Postal 89.000
Caracas, 1080-A, Venezuela

Elías Moreno
Departamento de Estadistica
Universidad de Grandada
Granada, España

On the Robustness of the Intrinsic Bayes Factor for Nested Models

discussion by
WALTER RACUGNO
University of Cagliari

Robustness analyses for the evidence of model M_1 versus M_2, with respect to several inputs is certainly a relevant topic in Bayesian inference.

Sansò, Pericchi and Moreno approach the argument using as a measure of evidence the intrinsic Bayes factor (IBF), introduced in Berger and Pericchi (1993), to allow the use of improper priors; this guarantees, in a certain sense, an "objective" or automatic procedure through default priors.

The work appears as a remarkable effort to construct "inherently" robust situations in the comparison of nested models; in addition, it presents interesting ideas. For example, from a conceptual point of view, relating the robustness to the width of the class of priors (in the specific case intrinsic priors) producing Bayes factors asymptotically equivalent to arithmetic intrinsic Bayes factors (AIBF), and reversing ("dualizing"), the standard attitude towards robustness is very interesting.

But, here too, the whole building stands on the "quicksand" of improper priors and the IBF itself is not sufficient to rid the Bayesian inference of all the problems introduced in the procedure by impropriety.

To support my point of view I am going to show an example, by F. Bertolino and myself (1995), and I will present a few other points about the paper.

A. Let $X = (X_1, \ldots, X_n)$ denote a sample and suppose the X_i are i.i.d. $Po(\lambda), \lambda \in R^+$, under M_1, and i.i.d. NegBin (ν, θ), with $\nu = 1$ and $\theta \in (0,1)$, under M_2.

Assume, respectively, the priors

$$\pi_1(\lambda) = c_1 \lambda^{-\alpha} \text{ and } \pi_2(\theta) = c_2(\theta(1-\theta))^{-\beta},$$

with $\alpha, \beta \in [0,1]$ and c_1, c_2 unspecified constants.

The AIBF is defined by

$$B_{21}^N = B_{21}^N \frac{1}{L} \sum_{\ell=1}^{L} B_{12}^N(x(\ell))$$

where

$$B_{21}^N = \frac{c_2 B(n-\beta+1, t-\beta+1)}{c_1 \prod_i (1/x_i!) \Gamma(t-\alpha+1)/n^{t-\alpha+1}}, \; (t = \Sigma x_i),$$

and for a minimal training sample of size $k = 1$

$$B_{12}^N(x(\ell)) = \frac{c_1(1/x(\ell)!)\Gamma(x(\ell)-\alpha+1)}{c_2 B(2-\beta, x(\ell)-\beta+1)}.$$

Let us consider the data sets $x^{(1)} = \{1,1,1,1,1,1,1,1,2,2\}$, $x^{(2)} = \{0,0,1,1,1,1,1,1,2,2\}$, $x^{(3)} = \{0,0,0,0,0,1,1,2,3,5\}$ and $x^{(4)} = \{0,0,0,0,0,0,0,0,0,1\}$. For different values of α and β, we have

Table 1

α	β	$B_{21}^{AI}(x^{(1)})$	$B_{21}^{AI}(x^{(2)})$	$B_{21}^{AI}(x^{(3)})$	$B_{21}^{AI}(x^{(4)})$
0	0	.040	.182	19.404	1.818
1	0	.040	.171	16.041	.455
1	.01	.040	.170	15.992	.457
.99	0	.040	1.267	248.890	14.001
.99	.01	.040	1.260	247.500	14.039
1	1	.028	.119	10.228	1.000

When $\alpha = \beta = 0$, the conclusions appear comparatively sensible: for $x^{(1)}$ and $x^{(2)}, B_{21}^N$ gives more evidence to M_1 w.r.t. M_2 and vice versa for $x^{(3)}$, $x^{(4)}$. But when we assume α or β equal 1, $B_{12}^N(x(\ell))$ is not determined for $x(\ell) = 0$. In this case the AIBF can assume very discordant values, depending on the choice of α and β. The values in table 1 show that the instability of $B_{21}^{A1}(x)$ increases as the number of zeros in the sample is large.

B. As is known, $\bar{B}_{12} = \frac{1}{L}\sum_{\ell=1}^{L} B_{12}^N(x(\ell))$ can be extremely unstable, particularly in nonnested situations. Without discussing the merit of the potential adhoc solutions (such as geometric average, α-trimming or the elimination of particular observations), let us consider an example used by Berger and Pericchi (1993 p. 22) showing the origin of instability.

Two models are entertained for the real variable $X, M_1 : \exp(\theta)$ and $M_2 : \log N(\mu,\sigma)$. For M_1 and M_2 the standard noninformative priors $\pi_1(\theta) \propto \theta^{-1}$ and $\pi_2(\mu,\sigma) \propto \sigma^{-1}$ are assumed, respectively. For a sample set $X = (X_1,\ldots,X_n)$ and for $k = 2$, the predictives m_1^N and m_2^N in $B_{12}^N(x(\ell))$ are

$$m_1^N(x(\ell)) = (x_i + x_j)^{-2},\ m_2^N(x(\ell)) = (2x_i x_j \mid \log(x_i/x_j) \mid)^{-1}.$$

Problems arise when data rounding causes two observations to be equal or very close. Rather than try and avoid these difficulties by specific solutions, it could be useful to study the problem in the following way: besides the sample $x = (x_1,\ldots,x_n)$ let us consider a "similar" sample $x' = (x'_1,\ldots,x'_n)$, where $x'_i \in (x_i - \varepsilon, x_i + \varepsilon)$. The Bayes factor $B_{21}(x)$, calculated assuming proper priors, must be very close to the Bayes factor $B_{21}(x')$ with the same priors, since when ε is sufficiently small the two samples are practically equal. If we now consider two sequences of proper priors $\pi^{(\alpha)}(\theta)$ and $\pi_2^{(\beta)}(\sigma)$ that tend to the impropers $\pi_1(\theta)$ and $\pi_2(\sigma)$, if B_{21} is stable for x and x' (as is easily expected), then the instability of other measures of evidence is not acceptable.

C. By means of a specific example, Sansò, Pericchi and Moreno suggest an expected intrinsic Bayes factor (B_{21}^{EA1}) in situations in which the Bayes factor (B_{21}) does not have the wanted robustness w.r.t. a prior class. They consider $M_1 : N(0,1)$ and $M_2 : N(\theta, 1)$ and assume $\pi_2(\theta) = N(0, \tau^2)$. The resulting Bayes factor to compare M_1 with M_2 for a sample $X = (X_1, \ldots, X_n)$ is

$$B_{21} = (n\tau^2 + 1)^{-1/2} \exp\left(\frac{\bar{x}^2 n^2 \tau^2}{2(n\tau^2 + 1)}\right)$$

which, as is seen, lacks robustness w.r.t. τ^2. In order to have a more robust measure of evidence S.P. and M. consider an immaginary training sample of size $k = 1, X \sim N(\theta, 1)$, and obtain

$$B_{21}^{EAI} = (\tau^2 + 1)\left[(2\tau^2 + 1)(n\tau^2 + 1)\right]^{-1/2} \exp\left(\frac{\bar{x}^2 \tau^2 (n^2 \tau^2 (2n - 1) + n^2 - 1)}{2(n\tau^2 + 1)(2\tau^2 + 1)}\right).$$

In the case $n = 10$, $\bar{x} = 0.1$, the behavior of B_{21} and B_{21}^{EAI} for $\tau^2 \geq 1$ is illustrated in figure 1 of the paper, which shows the greater robustness of B_{21}^{EAI} as compared to B_{21}.

If we now consider the ratio $r = B_{21}/B_{21}^{EAI}$ for τ^2 and n fixed, we observe that it varies in a significant way with $\bar{x}$. This shows a different sensitivity of B_{21} and B_{21}^{EAI} to the data. Moreover for $n \to \infty$, B_{21}^{EAI} increases more rapidly than B_{21}, sending r to zero.

If it is true that B_{21}^{EA1} is more robust than B_{21} w.r.t. τ^2, it is also true that it can lead to very discordant inferential conclusions.

I would not like the analysis of robustness to be used as a Procrustean bed: we cannot cut its feet off to make it fit.

REFERENCES

Berger, J.O. and Pericchi, L.R. (1993). The intrinsic Bayes factor for model selection and prediction. Tech. Rep. 93-43C. Department of Statistics, Purdue University.

Bertolino, F. and Racugno, W. (1995) Is the intrinsic Bayes factor intrinsic? (submitted)

REJOINDER

B. SANSÓ, L. R. PERICCHI AND E. MORENO

The discussant considers three different aspects of the robustness of the IBF and illustrates them with some very provocative examples that are used to put the IBF methodology under test. It appears that the IBF performs quite well, however.

The discussant begin by insisting in the use of proper priors. Then the IBF, offers him a **method** of assessment, the intrinsic priors. To our knowledge this is the first general method of assessment of (conditionally) proper priors for model selection. This is particularly relevant in his second example. Berger and Pericchi (1993) calculated a conditionally proper intrinsic prior in this situation, and using it the problem dissapears. Alternatively, use of the Expected IBF or trimming 'almost' singular training samples would work as well. It is true that this potential instability has to be dealt with in practice, but we have offered different ways out of it.

His first example, is very challenging to any method of inference and the IBF methodology performs quite well: two separate discrete models are compared in absence of any prior information, and the sample information is just 10 observations. Two quite different alternative improper priors are considered. The sensitivity of the answers with respect to the parameters in the prior is mild for $x^{(1)}$ and dramatic for $x^{(4)}$, with the other cases somewhere in between. It could not be otherwise since the data in $x^{(4)}$ provide only one training sample and all the data are zero except for the last. It would be surprising that a sensible estimation of the parameters could be done, let alone model discrimination. The conclusion is that no robustness can be expected from a model comparison method when the information contained in the data is almost nil.

His third example is related to the method proposed in section 3 that consists of modifying a class of priors to yield a more robust Bayes factor in hypothesis testing problems. The feature pointed out by the discussant is that, given a specific prior, the 'robustified' Bayes factor tends to favour the alternative hypothesis quicker that the original Bayes factor, as the data gives more evidence against the null. Why is it an undesirable feature for the decision rule to react more quickly when the data show conflicting evidence with the null hypothesis? It appears that our method of section 3 is both more robust and more powerful.