L_1-Statistical Procedures and Related Topics
IMS Lecture Notes – Monograph Series (1997) Volume 31

Bayes factors for intrinsic and fractional priors in nested models. Bayesian robustness

Elias Moreno

University of Granada, Spain

Abstract: For model selection the Bayes factor is not well defined when using default priors since they are typically improper. To overcome this problem two methods have recently been proposed. These methods, intrinsic and fractional, are studied here as methods to producing proper prior distributions for model selection from the improper conventional priors for estimation. For nested models, fractional priors are here defined and a comparison with intrinsic priors introduced by Berger and Pericchi is carried out. Robustness of the Bayes factor as the prior varies over the classes of intrinsic and fractional priors, is studied. Some illustrative examples are provided.

Key words: Bayes factor, bayesian robustness, fractional priors, intrinsic priors, model selection.

AMS subject classification: Primary 62A15; secondary 62F15.

1 Introduction

Suppose that two models M_1 and M_2 are proposed to describe the data $z = (x_1, x_2, ..., x_n)$. Under model M_i the data are distributed as $f_i(z|\theta_i)$, and the prior distribution for θ_i is $\pi_i(\theta_i)$, $i = 1, 2$. The Bayesian way to compare the two models consists in computing the posterior odds

$$\frac{\Pr(M_2|z)}{\Pr(M_1|z)} = B_{21}(z)\frac{\Pr(M_2)}{\Pr(M_1)}.$$

Thus, the Bayes factor $B_{21}(z)$ encapsulates all what the data have to say about such a comparison. This Bayes factor is given by

$$B_{21}(z) = \frac{\int_{\Theta_2} f_2(z|\theta_2)\pi_2(\theta_2)d\theta_2}{\int_{\Theta_1} f_1(z|\theta_1)\pi_1(\theta_1)d\theta_1}.$$

To subjectively elicit the priors $\pi_i(\theta_i)$, $i = 1, 2$, is most cases a very difficult task. A way to alleviate this task is to elicit instead of a single prior for each model, a class of prior distributions $\Gamma = \{\pi_1(\theta_1), \pi_2(\theta_2)\}$ that maintains the features of the priors on which we are confident. Hence, the Bayes factors as the prior ranges over Γ takes now values in the range

$$(\inf_{\Gamma} B_{21}(z), \ \sup_{\Gamma} B_{21}(z)).$$

This range is generally too large and it typically gives $\inf_\Gamma B_{21}(z) = 0$, so that there are priors in Γ favouring M_1 and also prior favouring M_2. Hence, it does not allow to decide which of the model is supported by the data.

Another way to deal with the problem of model selection is to set as $\pi_i(\theta_i)$ the conventional prior for estimation of θ_i, say $\pi_i^N(\theta_i)$, which typically is improper, that is the integral $\int_{\Theta_i} \pi_i^N(\theta_i)d\theta_i$ diverges. This means that no normalization of $\pi_i^N(\theta_i)$ is possible so that it is defined up to an arbitrary multiplicative constant. This implies that $B_{21}(z)$ is defined up to a ratio of unspecified constants.

There are in the literature several ways either to specify the constants or to remove them from the analysis, see Akaike (1973), Schwarz (1978), Spiegelhalter and Smith (1982), O'Hagan (1995), Berger and Pericchi (1995, 1996), among others. For a recent review, see Kass and Raftery (1995).

In this paper we focus on intrinsic and fractional methodologies as methods to producing proper prior distributions for model comparison. This is motivated by the fact that while there are methods to produce prior for estimation that work reasonably well, there is a lack of such a methods for model selection.

Let us briefly summarize the intrinsic and fractional Bayes factors. The *intrinsic Bayes factor* (IBF) was proposed by Berger and Pericchi (1995, 1996). This is a partial Bayes factor based on a minimum training sample, say $x(l)$, which is a minimal subsample of the sample z such that $0 < \int_{\Theta_i} f_i(x(l)|\theta_i)\pi_i^N(\theta_i)d\theta_i < \infty$, $i = 1, 2$. This part of the sample is devoted to convert $\pi_i^N(\theta_i)$ into $\pi_i^N(\theta_i|x(l))$, which is now proper, and the rest of the data is devoted to construct the intrinsic Bayes factor for model comparison, using the $\pi_i^N(\theta_i|x(l))$ as priors. Thus the partial Bayes factor is defined as

$$B_{21}^I(x(-(l)|x(l))) = B_{21}^N(z) \ B_{12}^N(x(l))$$

where $B_{21}^N(z)$ is the Bayes factor for the improper priors $\pi_i^N(\theta_i)$, $i = 1, 2$, and sample z. Note that $B_{21}^I(x(-(l)|x(l)))$ does not depend on the arbitrary constant involved in the improper priors $\pi_i^N(\theta_i)$, $i = 1, 2$. To avoid the dependence of the partial Bayes factor on the particular $x(l)$, they introduced an average on the set of all training samples. Thus, the *arithmetic intrinsic* Bayes factor is defined as

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

where L is the number of training sample contained in z.

The fractional method, proposed by O'Hagan, considers the *fractional Bayes factor* (FBF) which, as the author motivates, *is defined by analogy with the partial Bayes factor to avoid the arbitrariness of choosing a particular training sample.* The FBF is defined as

$$B_{21}^b(z) = B_{21}^N(z)\,\frac{\int_{\Theta_1} f_1(z|\theta_1)^b \pi_1^N(\theta_1)d\theta_1}{\int_{\Theta_2} f_2(z|\theta_2)^b \pi_2^N(\theta_2)d\theta_2} \tag{2}$$

where b is a constant that depends on the sample size n. Notice that $B_{21}^b(z)$ does not depend on the arbitrary constants involved in the improper priors.

Both, the IBF and the FBF, contain $B_{21}^N(z)$ as a common factor. The other factors appearing in the right hand side of (1) and (2) can be considered as the correction term of $B_{21}^N(z)$ to avoid the dependence of the unspecified constants. These correction terms are different. Furthermore, while the IBF correction term is completely specified, the FBF correction term depends on b that has to be assessed. We will go back on this topic in Subsection 2.4.

A crucial property of the intrinsic method is that it is capable to generating proper priors. These priors are derived by imposing that the IBF is asymptotically equivalent to an actual Bayes factor for the so-called *intrinsic priors* (see Berger and Pericchi, 1996, for the definition of *intrinsic priors* and Moreno, Bertolino and Racugno, 1996, for a characterization). In some sense this guarantee that the IBF is a truly Bayes factor.

The question is if something similar can be said on the FBF. In Subsection 2.2 we produce a functional equation to derive *fractional priors* that enjoy the same spirit of the *intrinsic priors.* The solution of this equation is also discussed. Computation of the Bayes factor for intrinsic priors entails a robustness issue since they are not unique, see Subsection 2.1. It will be also shown that the solution to the fractional equation is not unique so that a similar robustness issue appears here. Robustness is studied in Subsection 2.3. In Section 3 some illustrative examples are given. Section 4 contains some conclusions.

2 The intrinsic and fractional priors for nested models

Suppose that the two sampling models under comparison $\{f_1(x|\theta_1), \theta_1 \in \Theta_1\}$, $\{f_2(x|\theta_2), \theta_2 \in \Theta_2\}$ are nested. This means that the following conditions are satisfied,

(i) $\Theta_1 \subset \Theta_2$,
(ii) $f_2(x|\theta_2) \equiv f_1(x|\theta_1)$, for $\theta_2 = \theta_1$.

Let $\pi_i^N(\theta_i)$, $i = 1, 2$, the improper priors chosen.

2.1 The intrinsic priors

The *intrinsic priors* are shown to be, see Moreno, Bertolino and Racugno (1996). any pair $(\pi_1(\theta_1), \pi_2(\theta_2))$ such that,

(a) $\pi_1(\theta_1)$ is any prior in the class

$$\Gamma_1 = \{\pi_1(\theta_1) : \int_{\Theta_1} \pi_1(\theta_1) d\theta_1 = 1, \ \int_{\Theta_2} T(\theta_2)\ \pi_1(\psi_1(\theta_2)) d\theta_2 = 1\}$$

where $T(\theta_2) = \frac{\pi_2^N(\theta_2)}{\pi_1^N(\psi_1(\theta_2))} E^{M_2}_{x(l)|\theta_2} B^N_{12}(x(l))$ and $\psi_1(\theta_2)$ is the limit point of the MLE $\hat{\theta}_1(z)$ for parameter θ_1 in M_1 when sampling from M_2 at point θ_2.

(b) For each $\pi_1(\theta_1) \in \Gamma_1$, $\pi_2(\theta_2)$ is given by

$$\pi_2(\theta_2) = T(\theta_2)\pi_1(\psi_1(\theta_2)).$$

For studying robustness of the Bayes factor with respect to intrinsic priors it is convenient to express Γ_1 as

$$\Gamma_1 = \{\pi_1(\theta_1) : \int_{\Theta_1} \pi_1(\theta_1) d\theta_1 = 1, \ \int_{\Theta_1} V(\theta_1)\ \pi_1(\theta_1) d\theta_1 = 1\},$$

where $V(\theta_1) = T(\theta_1) + \{\int_{\Theta_2(\theta_1)} T(\theta_2) d\theta_2\} 1_{\Theta_1^*}(\theta_1)$, $\Theta_2(\theta_1) = \{\theta_2 \in (\Theta_2 - \Theta_1) : \psi_1(\theta_2) = \theta_1\}$, the ψ_1-coset of θ_1 in $\Theta_2 - \Theta_1$, and Θ_1^* is the ψ_1-image of $\Theta_2 - \Theta_1$.

2.2 The *fractional priors*

Let us first precisely state what we mean by *fractional priors.*

Definition 1 *For the sampling models $\{f_i(x|\theta_i), i = 1, 2\}$ the proper priors $\pi_1(\theta_1), \pi_2(\theta_2)$ are called fractional priors if its Bayes factor and the FBF for $\pi_i^N(\theta_i)$, $i = 1, 2$, are asymptotically equivalent for some sequence $\{b_n\}$.*

The following condition is necessary in order to propose the fractional equation from which the fractional priors are derived.

Assumption (A) The nested models $\{f_i(x|\theta_i),\ \pi_i^N(\theta_i),\ i=1,2\}$ satisfy Assumption (A) if for some sequence $\{b_n\}$ the limit in probability $[P_{\theta_2}]$ of the correction term of the FBF is a degenerated random variable. In other words, under Assumption (A) there exists a function $F_{12}^{M_2}(\theta_2)$, which could be a constant, such that

$$F_{12}^{M_2}(\theta_2) = \lim_{n\to\infty}[P_{\theta_2}]\frac{\int_{\Theta_1} f_1(z|\theta_1)^{b_n}\pi_1^N(\theta_1)d\theta_1}{\int_{\Theta_2} f_2(z|\theta_2)^{b_n}\pi_2^N(\theta_2)d\theta_2}$$

For simplicity in notation the dependence of $F_{12}^{M_2}(\theta_2)$ on the sequence $\{b_n\}$ is not made explicitly.

Theorem 1 *Under Assumption (A), the fractional priors ($\pi_1(\theta_1)$, $\pi_2(\theta_2)$) are the solutions to the functional equation*

$$F_{12}^{M_2}(\theta_2) = \frac{\pi_2(\theta_2)}{\pi_2^N(\theta_2)}\,\frac{\pi_1^N(\psi_1(\theta_2))}{\pi_1(\psi_1(\theta_2))}. \tag{3}$$

Proof: If we expand $\frac{\pi_i(\theta_i)}{\pi_i^N(\theta_i)}$ around the MLE $\hat{\theta}_i(z)$, the Bayes factor for the fractional priors $\pi_1(\theta_1)$, $\pi_2(\theta_2)$, can be approximated as

$$\begin{aligned} B_{21}(z) &= \frac{\int_{\Theta_2} f_2(z|\theta_2)\pi_2^N(\theta_2)\frac{\pi_2(\theta_2)}{\pi_2^N(\theta_2)}d\theta_2}{\int_{\Theta_1} f_1(z|\theta_1)\pi_1^N(\theta_1)\frac{\pi_1(\theta_1)}{\pi_1^N(\theta_1)}d\theta_1} \\ &= B_{21}^N(z)\frac{\pi_2(\hat{\theta}_2(z))}{\pi_2^N(\hat{\theta}_2(z))}\,\frac{\pi_1^N(\hat{\theta}_1(z))}{\pi_1(\hat{\theta}_1(z))}(1+o(1)). \end{aligned}$$

Equating the limit in probability $[P_{\theta_2}]$ of the fractional Bayes factor given in (2) with the limit in probability $[P_{\theta_2}]$ of the above expression, we obtain

$$F_{12}^{M_2}(\theta_2) = \lim_{n\to\infty}[P_{\theta_2}]\frac{\pi_2(\hat{\theta}_2(z))}{\pi_2^N(\hat{\theta}_2(z))}\,\frac{\pi_1^N(\hat{\theta}_1(z))}{\pi_1(\hat{\theta}_1(z))},$$

where the left hand side follows from Assumption (A). Notice that we do not need to take limit in probability under model M_1 since it is nested in M_2. This gives (3) and proves the assertion. □

We remark that the fractional priors does not depends on the arbitrary constants involved in the improper priors. They cancel out in expression (3).

Corollary 1 *Fractional priors are any pair* $(\pi_1(\theta_1), \pi_2(\theta_2))$, *where* $\pi_1(\theta_1)$ *is any member of the class*

$$\Gamma_2 = \{\pi_1(\theta_1) : \int_{\Theta_1} \pi_1(\theta_1)d\theta_1 = 1, \int_{\Theta_2} S(\theta_2)\pi_1(\psi_1(\theta_2))d\theta_2 = 1\},$$

with

$$S(\theta_2) = F_{12}^{M_2}(\theta_2)\ \frac{\pi_2^N(\theta_2)}{\pi_1^N(\psi_1(\theta_2))},$$

and for each $\pi_1(\theta_1) \in \Gamma_2$,

$$\pi_2(\theta_2) = S(\theta_2)\pi_1(\psi_1(\theta_2)).$$

Proof: Equation (3) can be written as

$$\pi_2(\theta_2) = S(\theta_2)\pi_1(\psi_1(\theta_2)),$$

where $\pi_1(\theta_1)$ have to be a probability distribution such that $\pi_2(\theta_2)$ be also a probability distribution. This proves Corollary 1. □

It is convenient to rewrite class Γ_2 as

$$\Gamma_2 = \{\pi_1(\theta_1) : \int_{\Theta_1} \pi_1(\theta_1)d\theta_1 = 1, \int_{\Theta_1} H(\theta_1)\pi_1(\theta_1)d\theta_1 = 1\},$$

where

$$H(\theta_1) = S(\theta_1) + \left\{\int_{\Theta_2(\theta_1)} S(\theta_2)d\theta_2\right\} 1_{\Theta^*(\theta_1)}(\theta_1).$$

2.3 Robustness of the Bayes factor for intrinsic and fractional priors

From Subsection 2.1 it follows that the Bayes factor for the *intrinsic priors* $(\pi_1(\theta_1), \pi_2(\theta_2))$ is

$$B_{21}(z) = \frac{\int_{\Theta_2} f_2(z|\theta_2)T(\theta_2)\ \pi_1(\psi_1(\theta_2))d\theta_2}{\int_{\Theta_1} f_1(z|\theta_1)\pi_1(\theta_1)d\theta_1},$$

which is written only in term of $\pi_1(\theta_1)$.

It is easily shown that $B_{21}(z)$, can be expressed as

$$B_{21}(z) = \frac{\int_{\Theta_1} W(z;\theta_1)\ \pi_1(\theta_1)d\theta_1}{\int_{\Theta_1} f_1(z|\theta_1)\ \pi_1(\theta_1)d\theta_1}$$

where $\pi_1(\theta_1) \in \Gamma_1$ and

$$W(z;\theta_1) = f_2(z|\theta_1)T(\theta_1) + \{\int_{\Theta_2(\theta_1)} f_2(z|\theta_2)T(\theta_2)d\theta_2\}1_{\Theta_1^*}(\theta_1).$$

On the other hand, from Corollary 1 it follows that the Bayes factor for fractional priors $\tilde{B}_{21}(z)$ can be written only in term of $\pi_1(\theta_1) \in \Gamma_2$ as

$$\tilde{B}_{21}(z) = \frac{\int_{\Theta_2} f_2(z|\theta_2) S(\theta_2) \pi_1(\psi_1(\theta_2)) d\theta_2}{\int_{\Theta_1} f_1(z|\theta_1) \pi_1(\theta_1) d\theta_1}.$$

It is straightforward to show that this Bayes factor associated to fractional priors can be written as

$$\tilde{B}_{21}(z) = \frac{\int_{\Theta_1} G(z;\theta_1) \pi_1(\theta_1) d\theta_1}{\int_{\Theta_1} f_1(z|\theta_1) \pi_1(\theta_1) d\theta_1},$$

where

$$G(z;\theta_1) = f_2(z|\theta_1) S(\theta_1) + \left\{ \int_{\Theta_2(\theta_1)} f_2(z|\theta_2) S(\theta_2) d\theta_2 \right\} 1_{\Theta^*(\theta_1)}(\theta_1).$$

Global robustness of the Bayes factors $B_{21}(z)$ and $\tilde{B}_{21}(z)$ as π_1 ranges over Γ_1 and Γ_2 respectively, can be established by computing the ranges $(\inf_{\pi_1 \in \Gamma_1} B_{21}(z), \sup_{\pi_1 \in \Gamma_1} B_{21}(z)), (\inf_{\pi_1 \in \Gamma_2} \tilde{B}_{21}(z), \sup_{\pi_1 \in \Gamma_2} \tilde{B}_{21}(z))$. This involve a moment problem for which Theorem 2 summarizes the solution.

Theorem 2 *The infimum of the fractional Bayes factor as the priors range over the class of priors* Γ_2, *say* $\underline{\lambda} = \inf_{\pi_1 \in \Gamma_2} \tilde{B}_{21}(z)$, *is the unique solution in* $\underline{\lambda}$ *to the equation*

$$\sup_{d \in R} \inf_{\theta_1 \in \Theta_1} [G(z;\theta_1) - \underline{\lambda} f_1(z|\theta_1) + d(1 - H(\theta_1))] = 0.$$

The sup *is obtained by interchanging in the above expression* sup *with* inf. *A similar statement can be given for the Bayes factor for the intrinsic priors.*

Proof: The proof follows by using the linearization algorithm, see for instance Lavine, Wasserman and Wolpert (1993), and the so-called Generalized Moment Theory, Kemperman (1987), Salinetti (1994) and Liseo, Moreno and Salinetti (1996).

For data z, robustness of a Bayes factor $B_{21}(z)$ as the priors range over a given class Γ, is strictly achieved if either $\sup_{\pi \in \Gamma} B_{21}(z) < 1$ favouring M_1, or $\inf_{\pi \in \Gamma} B_{21}(z) > 1$, in which case M_2 is favoured. With obvious adaptations of the suggestion by Jeffreys (see Kass and Raftery, 1995) we would take the following interpretation:

if $1 < \inf_{\pi \in \Gamma_1} B_{21}(z) < \sqrt{10}$, the evidence against M_1 is small,
if $\sqrt{10} < \inf_{\pi \in \Gamma_1} B_{21}(z) < 10$, the evidence against M_1 is substantial,
if $10 < \inf_{\pi_1 \in \Gamma_1} B_{21}(z) < 100$, the evidence against M_1 is strong and
if $100 < \inf_{\pi_1 \in \Gamma_1} B_{21}(z)$, the evidence against M_1 is decisive.

2.4 The role of the sequence $\{b_n\}$ *in the fractional priors*

An important difference between the intrinsic priors and the fractional priors is that while the former are derived automatically from the specification of the models $\{f_i(x|\theta_i), \pi_i^N(\theta_i), i = 1, 2\}$, the latter needs in addition to assess the sequence $\{b_n\}$. In fact, for producing fractional priors we already have a restriction on this sequence since Assumption (A) has to be satisfied. Nevertheless, this does not guarantee the uniqueness of the sequence, so that an additional convention has to be imposed. Let us illustrate the assertion with the following simple example.

Example 1 *Consider the nested models*

$$M_1 : f_1(x|\theta_1) = N(x|\theta_1, 1),\ \pi_1(\theta_1) = 1_{\{0\}}(\theta_1),$$
$$M_2 : f_2(x|\theta_2) = N(x|\theta_2, 1),\ \pi_2^N(\theta_2) \propto 1_R\ (\theta_2),$$

that is, we are testing that the mean of a normal distributions is 0 *versus it is different from* 0*. Notice that the prior for the first model is proper and the prior for the second is the conventional uniform improper prior.*

The intrinsic priors for this models can be shown to be the unique pair

$$\pi_1(\theta_1) = 1_{\{0\}}(\theta_1),\ \pi_2(\theta_2) = N(\theta_2|0, 2).$$

On the other hand the fractional priors are derived as follows. For a given sequence $\{b_n\}$, we have

$$\begin{aligned} F_{12}^{M_2}(\theta_2) &= \lim_{n\to\infty}[P_{\theta_2}]\frac{\int_{\Theta_1} f_1(z|\theta_1)^{b_n}\pi_1^N(\theta_1)d\theta_1}{\int_{\Theta_2} f_2(z|\theta_2)^{b_n}\pi_2^N(\theta_2)d\theta_2} \\ &= \lim_{n\to\infty}[P_{\theta_2}]\frac{\sqrt{nb_n}}{\sqrt{2\pi}}\exp\{-b_n\frac{n\bar{x}}{2}\}, \end{aligned}$$

where $\bar{x} = \frac{1}{n}\sum_1^n x_i$. The sequences $\{b_n\}$ proposed by O 'Hagan (1995) are

$$b_n = \frac{m_0}{n},\ m_0 = 1, 2, ..., n-1,$$
$$b_n^{'} = \frac{\sqrt{m_0}}{n},$$
$$b_n^{''} = \frac{\log n}{n}.$$

The sequences $\{b_n^{'}\}, \{b_n^{''}\}$ do not satisfy Assumption (A), so that they do not produce fractional priors. Therefore, we are left with sequences of the form $\{b_n = \frac{m_0}{n}, m_0 = 1, 2, ..., n-1\}$, for which we obtain

$$F_{12}^{M_2}(\theta_2) = \frac{\sqrt{m_o}}{\sqrt{2\pi}}\exp\{-m_0\frac{\theta_2^2}{2}\} = N(\theta_2|0, \frac{1}{m_0}).$$

Thus, the *fractional priors* are

$$\pi_1(\theta_1) = 1_{\{0\}}(\theta_1),\ \pi_2(\theta_2) = N(\theta_2|0, \frac{1}{m_0}),\ m_0 = 1, ..., n-1.$$

Note that $\pi_2(\theta_2)$ is not unique but depends on m_0. The convention we will take is to fix m_0 as the minimal training sample size (Berger and Mortera, 1995). In this case $m_0 = 1$. For this assessment the fractional prior for M_2 is the density $N(\theta_2|0,1)$.

3 Examples

Let us illustrate the behaviour of the Bayes factor for intrinsic and fractional priors for two standard problems. The first is one sided testing on the mean of a normal distribution and the second is a two sided testing.

Example 2 ***One sided testing.*** *Let X be a random variable $N(x|\theta,1)$ distributed. Suppose that we are interested in testing $H_1 : \theta \geq 0$ versus $H_2 : \theta < 0$.*

A formulation of this testing problem in a nested context would be to compare the two models

$$M_1 : f_1(x|\theta_1) = N(x|\theta_1, 1),\ \pi_1^N(\theta_1) \propto 1_{[0,\infty)}(\theta_1),$$
$$M_2 : f_2(x|\theta_2) = N(x|\theta_2, 1),\ \pi_2^N(\theta_2) \propto 1_{(-\infty,\infty)}(\theta_2),$$

where the priors are the standard improper prior for location parameter. it is easy to see that $\psi_1(\theta_2) = \theta_2 1_{[0,\infty)}(\theta_2)$.

The class of intrinsic priors is

$$\Gamma_1 = \{\pi_1 : \int_0^\infty \Phi(\frac{\theta_1}{\sqrt{2}})\pi_1(\theta_1)d\theta_1 = 1 - \frac{k}{\sqrt{\pi}}, 0 \leq k \leq \frac{1}{2}\sqrt{\pi}\},$$

where k is the value of $\pi_1(\theta_1)$ at the discontinuity point $\theta_1 = 0$ (see Moreno, Bertolino and Racugno, 1996) and $\Phi(\theta_1)$ is the standard cumulative distribution function at point θ_1. Thus, for a given sample $(\bar{x}, n)$ and $\pi_1(\theta_1) \in \Gamma$, the Bayes factor for M_2 against M_1 can be shown to be

$$B_{21}(\bar{x}, n) = \frac{kN(\bar{x}, n) + \int_0^\infty \exp\{-\frac{n(\bar{x}-\theta_1)^2}{2}\}\Phi(\frac{\theta_1}{\sqrt{2}})\pi_1(\theta_1)d\theta_1}{\int_0^\infty \exp\{-\frac{n(\bar{x}-\theta_1)^2}{2}\}\pi_1(\theta_1)d\theta_1},$$

where $N(\bar{x}, n) = \int_{-\infty}^0 \exp\{-\frac{n(\bar{x}-\theta_2)^2}{2}\}\Phi(\frac{\theta_2}{\sqrt{2}})d\theta_2$. In Moreno, Bertolino and Racugno (1966) it was shown that $\sup_{\pi_1 \in \Gamma_1} B_{21}(\bar{x}, n)$ is infinity for any sample point. Since for $\bar{x} \geq 0$ we have that $\inf_{\pi_1 \in \Gamma_1} B_{21}(\bar{x}, n)$ is less than 1 we conclude that the Bayes factor is not robust with respect to the intrinsic priors.

A limiting procedure was In Moreno, Bertolino and Racugno (1966) introduced to overcome the lack of robustness. This procedure is based on the fact that when the prior for the simple model M_1 is proper, then the intrinsic prior for the complex model always exists and it is unique.

The idea is to take the restriction of $\pi_1^N(\theta_1)$ on an increasing sequence $\{C_n\}$ of subsets of Θ_1 such that $\int_{C_n} \pi_1^N(\theta_1)d\theta_1 < \infty$ and $\lim_{n\to\infty} C_n = \Theta_1$. Then we construct the associated sequence of intrinsic priors for M_2 and we take the limit of the corresponding sequence of Bayes factors. Under rather general conditions the limit was proved to be independent on the particular sequence $\{C_n\}$ we have chosen.

Applying this procedure to our example the limiting Bayes factor turns out to be

$$B_{21}(\bar{x}, n) = \frac{\Phi(\frac{\bar{x}\sqrt{n}}{\sqrt{2n+1}})}{\Phi(\bar{x}\sqrt{n})}.$$

On the other hand, the fractional priors for this problems are the following. For sequences $\{b_n\}$ of the form $b_n^{'} = \frac{\sqrt{m_0}}{n}$, and $b_n^{''} = \frac{\log n}{n}$,

$$\begin{aligned} F_{12}^{M_2}(\theta_2) &= \lim_{n\to\infty}[P_{\theta_2}]\frac{\int_{\Theta_1} f_1(z|\theta_1)^{b_n}\pi_1^N(\theta_1)d\theta_1}{\int_{\Theta_2} f_2(z|\theta_2)^{b_n}\pi_2^N(\theta_2)d\theta_2} \\ &= 1_{[0,\infty)}(\theta_2), \end{aligned}$$

so that

$$\pi_2(\theta_2) = \pi_1(\theta_2)1_{[0,\infty)}(\theta_2),$$

where $\pi_1(\theta_1)$ is any probability density on $[0, \infty)$. Therefore, the Bayes factor for any data z and any fractional prior turns out to be

$$\tilde{B}_{21}(z) = 1.$$

Therefore, the above sequences $\{b_n^{'}\}$, $\{ b_n^{''}\}$ produce *fractional priors* that gives a non sensible Bayes factor.

If we choose $b_n = \frac{m_o}{n}$ with $m_o = 1$, the minimal training sample size, the class of fractional priors turns out to be

$$\pi_2(\theta_2) = \frac{k}{\sqrt{2\pi}}1_{(-\infty,0)}(\theta_2) + \Phi(\theta_2)\pi_1(\theta_2)1_{(0,\infty)}(\theta_2)$$

where π_1 is any prior in the class

$$\Gamma_2 = \{\pi_1 : \int_0^\infty \Phi(\theta_1)\pi_1(\theta_1)d\theta_1 = 1 - \frac{k}{\sqrt{2\pi}}, 0 \leq k \leq \sqrt{\frac{\pi}{2}}\},$$

k being the value of $\pi_1(\theta_1)$ at the discontinuity point $\theta_1 = 0$. The Bayes factor for $\pi_1 \in \Gamma_2$ is given as

$$\tilde{B}_{21}(\bar{x}, n) = \frac{k \int_{-\infty}^{0} \exp(-\frac{n(\bar{x}-\theta_2)^2}{2})\Phi(\theta_2)d\theta_2 + \int_0^\infty \exp(-\frac{n(\bar{x}-\theta_1)^2}{2})\Phi(\theta_1)\pi_1(\theta_1)d\theta_1}{\int_0^\infty \exp(-\frac{n(\bar{x}-\theta_1)^2}{2})\pi_1(\theta_1)d\theta_1}.$$

It can be seen that $\tilde{B}_{21}(\bar{x}, n)$ is not robust as the prior ranges over class Γ_2. To overcome this lack of robustness we can apply the same limiting procedure considered for the Bayes factor for intrinsic priors. A difficulty we find here is that it is not necessarily true that for a proper prior for the simple model and improper for the complex, the corresponding fractional prior for the complex model is proper. Indeed, if $\pi_1(\theta_1)$ is a proper prior for M_1 and $\pi_2^N(\theta_2)$ the improper prior for M_2, the corresponding fractional prior for M_2 is given as

$$\pi_2(\theta_2) = \pi_2^N(\theta_2) \lim_{n\to\infty} [P_{\theta_2}] \frac{\int_{\Theta_1} f_1(z|\theta_1)^{b_n} \pi_1(\theta_1)d\theta_1}{\int_{\Theta_2} f_2(z|\theta_2)^{b_n} \pi_2^N(\theta_2)d\theta_2},$$

which is not necessarily a probability density.

Fortunately, if in this example we take as $\pi_1(\theta_1) = \frac{1}{b-a} 1_{(a,b)}(\theta_1)$, the restriction of $\pi_1^N(\theta_1)$ to the interval (a, b), and $b_n = \frac{1}{n}$ then the corresponding fractional prior for model M_2 is

$$\pi_2(\theta_2) = \frac{1}{b-a}(\Phi(b-\theta_2) - \Phi(a-\theta_2)),$$

which is a probability density for any values of a and b. The Bayes factor for this fractional priors is

$$\tilde{B}_{21}^{(a,b)}(\bar{x}, n) = \frac{\int_{-\infty}^{\infty} \exp(-\frac{n(\bar{x}-\theta_2)^2}{2})(\Phi(b-\theta_2) - \Phi(a-\theta_2))\frac{1}{b-a}d\theta_2}{\int_0^\infty \exp(-\frac{n(\bar{x}-\theta_1)^2}{2})\frac{1}{b-a}d\theta_1}.$$

The limit when $a \to 0$ and $b \to \infty$ results

$$\tilde{B}_{21}(\bar{x}, n) = \lim_{n\to\infty} \tilde{B}_{21}^{(a,b)}(z) = \frac{\Phi(\frac{\bar{x}\sqrt{n}}{\sqrt{n+1}})}{\Phi(\bar{x}\sqrt{n})},$$

which is very close to the limiting Bayes factor for intrinsic priors.

We have already said that the intrinsic methodology always generates a class of proper priors distributions (intrinsic priors) for model selection when the models are nested. The fractional methodology, however, not necessarily generates proper priors. The following example illustrates this assertion.

Example 3 ***Two sided testing.*** *Suppose we have to choose between the following two nested models,*

$$M_1 : f_1(x|\theta_1) = N(x|0, \sigma_1^2),\ \pi_1^N(\theta_1) \propto \frac{1}{\sigma_1} 1_{(0,\infty)}(\sigma_1),$$
$$M_2 : f_2(x|\theta_2) = N(x|\mu, \sigma_2^2),\ \pi_2^N(\theta_2) \propto \frac{1}{\sigma_2^2} 1_{R\times(0,\infty)}\ (\mu, \sigma_2).$$

The parameter spaces are respectively $\Theta_1 = 0 \times R^+$ and $\Theta_2 = R \times R^+$ and the improper priors are given by the Jeffreys rule. It is easy to see that $\psi_1(\theta_2) = \sqrt{\mu^2 + \sigma_2^2}$.

For the data $z = (x_1, x_2, ..., x_n)$ and a given sequence $\{b_n\}$, some algebra shows that

$$\begin{aligned} F_{12}^{M_2}(\theta_2) &= \lim_{n\to\infty} [P_{\theta_2}] \frac{\int_{\Theta_1} f_1(z|\theta_1)^{b_n} \pi_1^N(\theta_1) d\theta_1}{\int_{\Theta_1} f_2(z|\theta_2)^{b_n} \pi_2^N(\theta_2) d\theta_2} \\ &= \lim_{n\to\infty} [P_{\theta_2}] \frac{\sqrt{nb_n}}{\sqrt{2\pi}} \left(\frac{s^2}{s^2 + \bar{x}^2} \right)^{b_n \frac{n}{2}}, \end{aligned}$$

where $s^2 = \frac{1}{n}\sum_1^n (x_i - \bar{x})^2$ and $\bar{x} = \frac{1}{n}\sum_1^n x_i$. For sequences of the form $\{b_n'\}$, $\{b_n''\}$, Assumption (A) is satisfied, but $F_{12}^{M_2}(\theta_2) = 0$ and consequently no proper prior is fractional.

For sequences of the form $\{b_n = \frac{m_0}{n},\ m_0 = 1, 2, ..., n-1\}$,

$$F_{12}^{M_2}(\theta_2) = \frac{\sqrt{m_0}}{\sqrt{2\pi}} \left(\frac{\sigma_2^2}{\sigma_2^2 + \mu^2} \right)^{m_0/2},$$

and thus the class of *fractional priors* is $(\pi_1(\theta_1),\ \pi_2(\theta_2))$, where

$$\pi_2(\theta_2) = \frac{\sqrt{m_0}}{\sqrt{2\pi}} \ \frac{\sigma_2^{m_0-2}}{(\sigma_2^2 + \mu^2)^{\frac{m_0-1}{2}}} \pi_1(\psi_1(\theta_2)),$$

and $\pi_1(\theta_1)$ is any prior in the class

$$\begin{aligned} \Gamma_2 &= \{\pi_1(\theta_1) : \int_0^\infty \pi_1(\sigma_1) d\sigma_1 \\ &= 1, \int_0^\infty \pi_1(\sigma_1) d\sigma_1 \frac{\sqrt{m_0}}{\sqrt{2\pi}} \int_0^\pi (\sin\varphi)^{m_0-2} d\varphi = 1\}. \end{aligned}$$

It is easy to see that this class is empty for any $m_0 \geq 1$. In particular for $m_0 = 2$, the minimal training sample size for this problem, we have

$$\Gamma_2 = \{\pi_1(\theta_1) : \int_0^\infty \pi_1(\sigma_1) d\sigma_1 = 1, \int_0^\infty \pi_1(\sigma_1) d\sigma_1 = \frac{1}{\sqrt{\pi}}\},$$

that is clearly empty.

Therefore, the sequences $\{b_n\}$ suggested in the literature of fractional methodology do not generate proper prior distributions for this two sided testing problem.

However, the class of intrinsic priors is

$$\Gamma_1 = \{\pi_1(\theta_1) : \int_{\Theta_1} \pi_1(\theta_1) d\theta_1 = 1, \ \int_{\Theta_1} \frac{1}{\sigma_1} \pi_1(\theta_1) d\theta_1 = 2(\sqrt{\pi} - 1)\},$$

that is not empty.

4 Conclusions

In this paper we have considered the intrinsic prior distributions (Berger and Pericchi, 1995 and 1996), and introduced the notion of fractional priors. This permits to focus the intrinsic and fractional methodologies as tools for generating proper prior distributions for model comparison from the conventional improper priors for estimation. The considered models have assumed to be nested and the main conclusions are:

The intrinsic priors always exist and form a class given by generalized moment constraints. The Bayes factor for intrinsic priors is not necessarily robust, but the limit intrinsic procedure (Moreno, Bertolino and Racugno, 1996) solves this lack of robustness.

The fractional methodology, however, not always generates proper priors (fractional priors). When fractional priors there exist, we have found that the sequence $\{b_n = \frac{m_0}{n}, n \geq 1\}$ with m_0 equal to the minimal training sample size, is the appropriate selection among those recommended by O'Hagan (1995). The associated Bayes factor is then very close to the Bayes factor for intrinsic priors. Furthermore, calculations are quite simple.

Extension of this theory to non-nested models is an interesting topic that deserves more research. It is a work in progress that will be formalized elsewhere.

References

[1] Akaike, H. (1983). Information measures and model selection. *Bull. Int. Statist. Inst.* **50**, 277-290.

[2] Berger, J.O. and Mortera, J. (1995). Discussion to A. O'Hagan (1995). *J. R. Statist. Soc.* B **57**, 99-138.

[3] Berger, J.O. and Pericchi, L.R. (1995). The intrinsic Bayes factor for linear models (with discussion). In *Bayesian Statistic,* Eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Vol. 4, pp. 35-60. Oxford University Press.

[4] Berger, J.O. and Pericchi, L.R. (1996). The intrinsic Bayes factor for model selection and prediction . *J. Am. Statist. Assoc.* **91**, 109-122.

[5] Jeffreys, H. (1961). *Theory of Probability.* London: Oxford University Press.

[6] Kass, R.E. and Raftery, A. (1995). Bayes factors. *J. Am. Statist. Assoc.* **90**, 773-795.

[7] Kemperman, J.H.B. (1987). Geometry of the moment problem. *Proc. of Symposia in Applied Mathematics* **37**, 16-53.

[8] Lavine, M., Wasserman, L. and Wolpert, R. (1993). Linearization of Bayesian robustness problems. *J. Am. Statist. Assoc.* **86**, 964-971.

[9] Liseo, B., Moreno, E. and Salinetti, G. (1996). Bayesian robustness for classes of priors with given marginals (with discussion). *Proc. of the Workshop on Bayesian Robustness*, Eds. Berger, J.O., Betro, B., Moreno, E., Pericchi, L.R., Ruggeri, F., Salinetti, G. and Wasserman, IMS Lectures Notes-Monograph Series, Vol. 29, pp. 103-120.

[10] Moreno, E., Bertolino F. and Ragcugno W. (1996). The intrinsic priors in model selection and hypotheses testing. Technical Report, University of Granada.

[11] O'Hagan, T. (1995). Fractional Bayes factor for model comparison (with discussion). *J. R. Statist. Soc.* B **57**, 99-138.

[12] Salinetti, G. (1994). Discussion on "An overview of Robust Bayesian Analysis" by J.O. Berger. *Test* **3**, 5-125.

[13] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461-464.

[14] Spiegelhalter, D.J. and Smith, A.F.M. (1982). Bayes factor for linear and log-linear models with vague prior information. *J. R. Statist. Soc.* B **44**, 377-387