Objective Priors for Model Selection in One-Way Random Effects Models

Gonzalo García-Donato
Universidad de Castilla-La Mancha
Spain

Dongchu Sun
University of Missouri
USA

May 10, 2005

Abstract

It is broadly accepted that the Bayes factor is a key tool in model selection. Nevertheless, it is an important, difficult and still open question which priors should be used to develop objective (or default) Bayes factors. We consider this problem in the context of the one-way random effects model. Arguments based on concepts like orthogonality, matching predictive, and invariance are used to justify a specific form of the priors, in which the (proper) prior for the new parameter (using Jeffreys’ terminology) has to be determined. Two different proposals for this proper prior have been derived: the intrinsic priors and the divergence based priors, a recently proposed methodology. It will be seen that the divergence based priors produce consistent Bayes factors. The methods are illustrated on examples and compared with other proposals. Finally, the divergence based priors and the associated Bayes factor are derived for the unbalanced case.

Keywords: Consistency; Divergence based priors; Intrinsic priors; Matching predictive priors; Objective Bayes factors; Orthogonality.

1 Introduction and notation

Consider the random effects model

\[ M : y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n, \]

where \( e_{ij} \sim N(0, \sigma^2) \), iid \( \forall i, j \) and \( a_i \sim N(0, \sigma_a^2) \), \( \forall i \), and \( k > 1 \). A modern revision as well as illustrations of this model can be found in Rencher (2000) and Ravishanker and Dey (2002). The variances \( \sigma_a^2 \) and \( \sigma^2 \) are often called variance components. It is well known (Hill, 1965; Box and Tiao, 1973) that the usual estimators of \( \sigma_a^2 \) can produce negative estimations, a major disadvantage of the classical methods.

*Address for correspondence: Gonzalo García-Donato, Department of Economy, Plaza Universidad 2, 02071 Albacete, Spain. Email:GonzaloGarciaDonato@uclm.es
The model above can be expressed in a more compact way as

\[ M : p(y \mid \mu, \sigma, \sigma_0) = \prod_{i=1}^{k} N_n(y_i \mid \mu 1_n, \Sigma), \]

where \( 1_n = (1, \ldots, 1)', \Sigma = \sigma^2 I_n + \sigma^2_0 J_n, \) \( I_n \) the identity matrix, \( y_i = (y_{i1}, \ldots, y_{in})' \) and \( y = (y_1, \ldots, y_k)' \). In this paper we deal with the model selection problem of no difference between groups, i.e.

\[ M_1 : p_1(y \mid \mu, \sigma) = p(y \mid \mu, \sigma, 0), \quad \text{vs.} \quad M_2 : p_2(y \mid \mu, \sigma, \sigma_a) = p(y \mid \mu, \sigma, \sigma_a). \] (1)

In Jeffreys’ (1961) terminology, \( \mu, \sigma \) are common parameters, while \( \sigma_a \) is a new parameter.

We denote SSA and SSE the sums of squares corresponding to the factor and the error

\[ SSA = n \sum_{i=1}^{k} (\bar{y}_i - \bar{y})^2, \quad SSE = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2, \]

where \( \bar{y}_i \) is the mean for treatment \( i \) and \( \bar{y} \) is the overall mean. We also denote

\[ SST = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y})^2 = SSA + SSE, \]

the total sum of squares.

The approach adopted to solve this model selection problem is Bayesian. See Berger and Pericchi (2001) for a detailed description of the advantages of Bayesian methods for model selection). Although there are several Bayesian alternatives to solve a model selection problem (see e.g. Bernardo, 1999; Gelfand and Ghosh, 1998; Ibrahim, Chen and Sinha, 2001; Spiegelhalter, Best and Carlin, 1998; Goutis and Robert, 1998). Our preferred one is the Bayes factor (see Kass and Raftery, 1995). The Bayes factor has a clear meaning in terms of “measure of evidence” in favour of a model or hypothesis (Berger and Delampady, 1987; Berger and Sellke, 1987). Further, the Bayes factor can be easily translated in posterior probabilities of models or hypotheses. Under the \( M \)-closed (Bernardo and Smith, 2000) approach (one of the models under competition is the true model), and assuming constant loss function, the optimal decision rule in a model selection problem is based on the Bayes factor (see Berger, 1985).

For problem (1), the Bayes factor in favor of \( M_1 \) and against \( M_2 \), \( B_{12} \), can be expressed as the quotient between the prior predictive marginal distributions evaluated at \( y \):

\[ B_{12} = \frac{m_1(y)}{m_2(y)}, \] (2)

where

\[ m_1(y) = \int p_1(y \mid \mu, \sigma) \pi_1(\mu, \sigma)d\mu d\sigma, \]

\[ m_2(y) = \int p_2(y \mid \mu, \sigma, \sigma_a) \pi_2(\mu, \sigma, \sigma_a)d\mu d\sigma d\sigma_a, \]

and \( \pi_1, \pi_2 \) are the corresponding prior distributions.
We assume that no other information external to the data and the competing models is available. In this scenario, known as “objective,” “minimum informative” or “by default,” is where the main difficulties with the Bayes factor appear. There are many situations in practice that demand a response to this problem: “Objective appearance,” “very weak prior information,” “early stages of the problem,” etc. In this situation, the crucial question to a Bayesian becomes what “objective” prior is suitable for a given model selection problem. Under our point of view, people should concentrate their efforts in proposing reasonable objective priors, instead of say, trying to reinvent “new” measures of evidence. This approach, adopted in this paper, is strongly motivated by the following principle:

**Principle (Berger and Pericchi, 2001):** Testing and model selection methods should correspond, in some sense, to actual Bayes factors, arising from reasonable prior distributions.

Except for notable exceptions (see Berger, Pericchi and Varshavsky, 1998), noninformative improper priors cannot be used, since the resulting Bayes factor is essentially arbitrary. As some authors have pointed out before, this fact makes evident the different nature of the “estimation problem” (the model is given) and the “model selection problem.” For clarity of exposition, in advance we use the term “noninformative” prior (denoted $\pi^N$) to refer to those priors introduced for estimation purposes (e.g. Jeffreys priors).

Starting from at least Jeffreys (1961), a number of different generic prior distributions for model selection have been proposed in the literature: the intrinsic priors (Berger and Pericchi, 1996a; Moreno, Bertolino and Racugno, 1998; O’Hagan 1995, 1997); the set of reference priors (Raftery, 1998); the unit information priors (Kass and Wasserman, 1995); the expected posterior priors (Perez and Berger, 2002) and in the normal context, the proposals of Jeffreys (1961) and Zellner and Siow (1980, 1984).

In this paper, objective prior distributions for the problem in (1) are derived. Our proposed priors are the result of a reasoned discussion about the prior for the common parameters $\mu, \sigma$ (Section 2) and the study of specific forms for the prior for the new parameter $\sigma_a$ (Section 3). For this last task, both the intrinsic priors (Berger and Pericchi, 1996a) and a novel methodology, based on a measure of divergence between the competing models will be used. Also, the consistency (asymptotically choosing the true model) of the associated Bayes factors will be analyzed. There are some proposals in the literature for this same problem: Westfall and Gonen (1996) and Pauler, Wakefield and Kass (1999). The prior proposed by Pauler, Wakefield and Kass (1999) is data-dependent, so in a strict sense it is not a genuine prior distribution and will not be considered in this paper. More attention will be given to the proposal by Westfall and Gonen (1996), which will be analyzed throughout the paper.

In Section 4, the different proposals are illustrated using real examples. In Section 5, extensions of the proposed priors to the unbalanced case will be addressed, and finally, the main conclusions of the paper will be summarized in Section 6.
2 Prior for common parameters

2.1 The proposal

We propose the following priors:

\[ \pi_1(\mu, \sigma) = \pi_N(\mu, \sigma), \quad \pi_2(\mu, \sigma, \sigma_a) = \pi_N(\mu, \sigma) \pi(\sigma_a | \mu, \sigma), \]  

where

\[ \pi_N(\mu, \sigma) = \sigma^{-1}, \quad \pi(\sigma_a | \mu, \sigma) = \sigma^{-1} f_n(\sigma_a/\sigma), \]  

and \( f_n \) (possibly dependent on \( n \)) is a proper density (i.e. integrates one over \( \mathbb{R}^+ \)). The common parameters are assumed to have the same noninformative prior distribution under \( M_1 \) and \( M_2 \). Note, this prior \( \pi_N \), is the reference prior (Berger and Bernardo, 1992a) for the simple model \( M_1 \). The new parameter, \( \sigma_a \), is assumed to have a proper prior distribution (hence avoiding the arbitrariness of Bayes factor) scaled by the standard deviation, \( \sigma \). Furthermore, \( \mu, \sigma \) are, a priori under \( M_2 \), independent of the ratio \( \rho = \sigma_a/\sigma \). The exact form of the function \( f_n \) will be further discussed in Section 3.

In order to properly justify the priors in (3), the following expression of the associated Bayes factor will prove to be useful.

**Proposition 1.** The Bayes factor, in favor of model \( M_2 \), for problem (1) produced with priors in (3) is

\[ B_{21} = \int_0^\infty \left(1 + nt^2\right)^{-(k-1)/2} \left(1 - \frac{nt^2}{1 + nt^2} \frac{SSA}{SST}\right)^{-(nk-1)/2} f_n(t) \, dt. \]  

Proof. See the Appendix.

\( B_{21} \) has really a simple expression and can be evaluated using an unidimensional integral. \( B_{21} \) depends on the usual sums of squares, which can be obtained using standard statistical packages.

2.2 Justification

In the recent literature, several important arguments have been used to decide whether objective priors for model selection are, in a broad sense, objectively reasonable. We use these criteria to judge the priors in (3) and conclude that the proposal is very reasonable.

- Matching predictive priors. Under this perspective (see Berger and Pericchi 1996b, 2001), when comparing models \( M_1 \) and \( M_2 \), the priors \( \pi_1, \pi_2 \) should be chosen so that the predictive distributions \( m_1(\tilde{y}) \) and \( m_2(\tilde{y}) \) are as close as possible if \( \tilde{y} \) is an imaginary sample of minimum size. This is equivalent of saying that \( B_{21} \) should be close to 1 in the situation of minimal sample information.
Interestingly, this idea was used by Jeffreys (1961) to propose objective priors to test about the mean in the normal model. As Jeffreys suggests (p. 269) “for \( n = 1 \) no decision would be reached in the absence of previous information.”

The minimal sample information in the problem (1) is obtained with \( k = 2, n = 1 \). It is immediate to show, using Proposition 1, that \( B_{21} = 1 \) under this situation. It has to be remarked that this property does not hold if another noninformative prior \( \pi^N \) is used or if \( f \) is not a proper density. Hence, it can be argued that the priors in (3) are well-calibrated.

- Orthogonality. We have used the term common parameters to refer to those parameters that appear in \( M_1 \) and \( M_2 \). Nevertheless, common parameters can change meaning from one model to another, and hence, using the same prior (even if it is noninformative) for common parameters does not seem to be reasonable. It is a standard practice (which again comes from at least Jeffreys, 1961) to identify, as same magnitudes, those common parameters orthogonal (i.e. the expected Fisher information matrix is diagonal) to the new parameters in \( M_2 \). This would justify the use of the same prior for orthogonal parameters. Besides, it is well-known (see Kass and Vaidyanathan, 1992) that under orthogonality, the Bayes factor is quite robust to the election of the (same) prior for common orthogonal parameters, justifying the use of noninformative priors.

It is easy to show that the expected Fisher information matrix for \( M_2 \) in (1) is

\[
\mathcal{I} = \begin{pmatrix}
\mathcal{I}_{\mu,\mu} & \mathcal{I}_{\mu,\sigma} & \mathcal{I}_{\mu,\sigma_a} \\
\mathcal{I}_{\mu,\sigma} & \mathcal{I}_{\sigma,\sigma} & \mathcal{I}_{\sigma,\sigma_a} \\
\mathcal{I}_{\mu,\sigma_a} & \mathcal{I}_{\sigma,\sigma_a} & \mathcal{I}_{\sigma_a,\sigma_a}
\end{pmatrix}, \text{ say,}
\]

with \( \mathcal{I}_{\mu,\sigma_a} = 0 \) and \( \mathcal{I}_{\sigma,\sigma_a} = O\left(\frac{1}{n}\right) \). Then, the common parameters \( \mu \) and \( \sigma \) are approximately (for moderate or large \( n \)) orthogonal to \( \sigma_a \).

- Invariance. It sometimes happens that using invariant arguments, a given model selection problem can be re-expressed in such a way that the common parameters (also called nuisance parameters) can be avoided. As noticed by Westfall and Gönen (1996) (based on the previous work by Westfall, 1989), for \( M_2 \) in (1), the maximal invariant statistic \( S \) under the composition (see the references above for details) of location shifts, scale changes and rotational invariant changes actions has a distribution \( p(s \mid \rho) \), which only depends on \( \rho = \sigma_a / \sigma \). If this invariant reduction is accepted, the problem (1) can be written as \( M_1^* : p_1(s) = p(s \mid 0) \) versus \( M_2^* : p_2(s \mid \rho) = p(s \mid \rho) \). Note, \( M_1^* \) nor \( M_2^* \) depend on nuisance parameters, and now only the prior \( \pi_2(\rho) \) has to be assigned. Interestingly, the Bayes factor in favor of \( M_2^* \) is

\[
B_{21}^* = \frac{m_2^*(s)}{m_1^*(s)} = \int_0^{\infty} (1 + n\rho^2)^{-(k-1)/2} \left(1 - \frac{n\rho^2}{1 + n\rho^2} \frac{SSA}{SST}\right)^{-(nk-1)/2} \pi(\rho) \, d\rho,
\]

that is, the same expression as that obtained in Proposition 1.

- Intrinsic priors. The intrinsic theory (Berger and Pericchi, 1996a; Moreno, Bertolino and Racugno, 1998) is a very intuitive and attractive procedure to obtain objective priors for
model selection. Broadly speaking, the intrinsic methodology converts non informative \( \pi_i^N \) (estimation) priors into priors \( \pi_i^I \) (called intrinsic priors) suitable for model selection. Asymptotically, the Bayes factor produced with \( \pi_i^I \) coincides with the intrinsic Bayes factor, an average of actual Bayes factors.

In Section 3, the intrinsic priors for problem (1) are derived. It will be seen that an improved version of these intrinsic priors \( \pi_i^1 \) and \( \pi_i^2 \) has the form proposed in (3), reinforcing hence the proposal.

### 3 Prior for New Parameter

In the previous section we have argued that the prior for \( M_2 \) should be of the form

\[
\pi_2(\mu, \sigma, \sigma_a) = \sigma^{-1} \pi(\sigma_a | \mu, \sigma), \quad \pi(\sigma_a | \mu, \sigma) = \sigma^{-1} f_n(\sigma_a / \sigma). \tag{7}
\]

We still have to deal with the difficult question of what \( f_n \) should be used. Apart from what we know that \( f_n \) has to be a proper density, its exact form is unknown. Westfall and Gönen (1996) propose priors that do not depend on the sample size \( n \), i.e. \( f_n = f \), finally proposing

\[
f_{WG}(t) = \frac{2t}{(1 + t^2)^2} \tag{8}
\]

(see Figure 1) as the result of an interesting discussion about the adequate form of the prior distribution. Besides, the authors prove the consistency of the associated Bayes factor when \( n \) (\( k \) fixed) or \( k \) (\( n \) fixed) grow.

Although specific proposals (like the one above) are valuable, we look for proposals that come from generic procedures and can be applied successfully to other scenarios. With this in mind, we explore several alternatives to assign \( f_n \). First, we use a novel methodology proposed by García-Donato (2003) under which the prior for the new parameter is a function of a positive measure of divergence between the competing models. Second, we derive the intrinsic priors. Several characteristics of the different proposal of \( f_n \) will be analyzed.

#### 3.1 Divergence based prior

The Divergence based (DB for short) priors are designed to provide reasonable objective proper prior distributions for the new parameter in a model selection problem.

The DB priors are defined as densities proportional to a positive measure of divergence between the models under comparison, raised to a negative power. To be more concise, for the problem (1) the proposal is

\[
\pi^D(\sigma_a | \mu, \sigma) \propto D_+(\mu, \sigma, \sigma_a)^{-q}, \quad q > q^*,
\]

where \( D_+(\cdot) \) is a positive measure of divergence (discrepancy) between \( M_1 \) and \( M_2 \) and \( q^* \geq 0 \), to be fixed is such that \( \pi^D \) is proper when \( q > q^* \). Note the objective nature of \( \pi^D \) because only intrinsic characteristics of the models are used in the definition.
The election of the specific measure $D_+\mu$ depends on the problem. But, as García-Donato (2003) points out, for regular problems, an attractive choice is using
\[ D_+(\mu, \sigma, \sigma_a) = 1 + \frac{D(\mu, \sigma, \sigma_a)}{nk}, \]
where
\[ D(\mu, \sigma, \sigma_a) = \int \log \frac{p_1(y | \mu, \sigma)}{p_2(y | \mu, \sigma, \sigma_a)} \{p_1(y | \mu, \sigma) - p_2(y | \mu, \sigma, \sigma_a)\} dy. \]
That is, $D$ is the (symmetrized) Kullback-Liebler (see Kullback, 1997) divergence between $M_1$ and $M_2$.

Dividing $D$ by $nk$ (the order of information, in $M_2$, as defined by Pauler, Wakefield and Kass, 1999), is a key part in the definition of DB prior, managing that the information incorporated into the prior is equivalent to the information contained in an imaginary sample of unitary size. The idea of using unitary sample information for priors in model selection has been proved to be successful for many authors (see Smith and Spiegelhalter, 1980; Kass and Wasserman, 1995).

Several reasons originally motivated the proposal (details can be found in García-Donato, 2003) of DB priors as a reasonable generic prior for new parameters:

- $\pi D$ is (under mild conditions) approximately Student, centered at the simple model and scaled by the block of the Fisher information matrix corresponding to the new parameter.

- In the normal scenario, the DB priors coincide with the proposals of Jeffreys (1961) and Zellner and Siow (1980, 1984) when $q$ is assigned to be $q^* + \frac{1}{2}$.

- The definition of DB does not depend on the dimension of the new parameter.

For irregular problems (e.g. parametric space dependent on observations), the convenience of other type of positive divergence (like those based on the minimum discrepancy) is discussed in García-Donato (2003). Finally, note that the DB priors are quite easy to obtain.

For the problem (1), straightforward algebra shows that
\[ D(\mu, \sigma, \sigma_a) = \frac{nk}{2\sigma^2} \left(\frac{n\sigma_a^4}{\sigma^2 + n\sigma_a^2}\right), \]
so the DB prior is
\[ \pi D(\sigma_a | \mu, \sigma) = c(n, q) \sigma^{-1} \left(1 + \frac{1}{2\left(\frac{\sigma}{\sigma_a}\right)^2 \left(\frac{\sigma_a}{\sigma}\right)^2 + n}\right)^{-q}, \quad q > q^* \] \hfill (9)
where $q^* = 1/2$ and
\[ c(n, q) = \left(\int_0^\infty \left(1 + \frac{ns^2}{2n + s^{-2}}\right)^{-q} ds\right)^{-1}. \] \hfill (10)
Note that $q > 1/2$ ensures the propriety of the prior distribution.
Interestingly, as $n$ becomes large, $\pi^D$ tends to be a (half) Student distribution with mode at zero:

$$\pi^D(\sigma_a \mid \mu, \sigma) \propto \left(1 + \frac{1}{2} \frac{\sigma_a^2}{\sigma^2}\right)^{-q}, \quad q > \frac{1}{2}, \quad (n \text{ large}).$$

The Student (and more concisely its particular case the Cauchy) density has been traditionally proposed as a sensible prior in model selection (see Jeffreys, 1961; Berger, Ghosh and Mukhopadhyay, 2003).

The DB prior is of the form (4) with

$$f^D_n(t) = c(n, q) \left(1 + \frac{1}{2} \frac{nt^4}{1 + nt^2}\right)^{-q}$$

(11) dependent on the number of replications $n$. We denote $B^D_{21}$ the Bayes factor in (5) with $f_n = f^D_n$. We now show that $B^D_{21}$ is consistent in a strong sense (almost surely) under the two more popular asymptotic scenarios:

- **Scenario I**: $k > 1$ fixed, $n \to \infty$;
- **Scenario II**: $n$ fixed, $k \to \infty$.

**Definition 1.** A Bayes factor $B_{21}$ is consistent under

- **Scenario I**, if as $n \to \infty$, then $B_{21} \to 0$, almost surely (a.s.) under $M_1$, and $B_{21} \to \infty$, a.s. under $M_2$.

- **Scenario II**, if as $k \to \infty$, then $B_{21} \to 0$, a.s. under $M_1$, and $B_{21} \to \infty$, a.s. under $M_2$.

**Theorem 1.** $B^D_{21}$ is consistent under Scenario I and Scenario II.

**Proof.** See the Appendix.

Originally, the parameter $q$ was introduced to ensure the propriety of the DB prior although $q$ can be used to accommodate desirable properties in the prior distribution. As $q$ is close to $q^*$ (1/2 in this case), the DB prior has flat tails; while as $q$ tends to be large, the DB prior is more peaked at zero. It is easy to show that if $\frac{1}{2} < q < 1$, then $\pi^D$ does not have moments (a property considered desirable by many, see Jeffreys, 1961; Liang, Paulo, Molina, Clyde and Berger, 2005). Nevertheless, our preferred election is $q = q^* + 1/2$ (so $q = 1$ for our problem). As noticed in García-Donato (2003), using $q = q^* + 1/2$ (recall that this assignment gives rise to the Jeffreys-Zellner-Siow priors in the linear models) seems to be a sensible election in many different scenarios. In the data sets analyzed in Section 4, we have used $q = 0.75, 1, 1.25$. The results obtained are quite robust to the election of $q$. For these three values of $q$, we have represented $f^D_t$, in Figure 2 for $n = 5$. 

8
3.2 Intrinsic prior

The intrinsic priors were originally introduced to provide a Bayesian justification to the intrinsic Bayes factors: asymptotically, the Bayes factor produced with the intrinsic prior, coincides with the intrinsic Bayes factor. Nowadays, the intrinsic priors are considered a device to produce sensible objective priors for model selection and have been derived, for specific problems, by many authors: Berger and Pericchi (1996b); Moreno, Torres and Casella (2005); Casella and Moreno (2002a); Casella and Moreno (2002b).

If \( \pi^N_1, \pi^N_2 \) are noninformative priors for \( M_1 \) and \( M_2 \) in (1), respectively, the intrinsic priors are defined as the priors \( \pi^I_1, \pi^I_2 \):

\[
\pi^I_1(\mu, \sigma) = \pi^N_1(\mu, \sigma), \quad \pi^I_2(\mu, \sigma, \sigma_a) = \pi^N_2(\mu, \sigma, \sigma_a) E^M_2(\mathcal{B}^N_{12}(y^*) | \mu, \sigma, \sigma_a),
\]

where \( m^N_i \) is the prior predictive under \( M_i \) obtained with respect to \( \pi^N_i \) and \( y^* \) is such that \( 0 < m^N_2(y^*) < \infty \) and there is no a subsample of \( y^* \) for which this inequality holds. \( y^* \) is usually called a minimal training sample.

To derive the intrinsic priors we use \( \pi^N_1(\mu, \sigma) = \sigma^{-1} \) (the reference prior for \( M_1 \)) and for \( M_2 \) the Jeffreys prior, \( \pi^{NJ}_2 \), and the reference prior, \( \pi^{NR}_2 \):

\[
\pi^{NJ}_2(\mu, \sigma, \sigma_a) = \frac{\sigma}{1 + nt^2} \Gamma(\frac{n-1}{2}) \left( \frac{n-1 + (1 + nt^2)^{-2}}{(1 + nt^2)^{-1/2}} \right)^{1/2},
\]

\[
\pi^{NR}_2(\mu, \sigma, \sigma_a) = \frac{\sigma}{1 + nt^2} \Gamma(\frac{n-1}{2}) \left( \frac{n-1 + (1 + nt^2)^{-2}}{(1 + nt^2)^{-1/2}} \right)^{1/2}.
\]

The reference priors were obtained by Berger and Bernardo (1992b). Some interesting properties related to these noninformative priors can be found in Ye (1995). It has to be noted that \( \pi^{NR}_2 \) is the reference prior for the ordered group \( \{ \sigma_a, (\sigma, \mu) \} \) which means that the parameters \( \sigma, \mu \) are of the same interest but less important than \( \sigma_a \). This is clearly the case in our problem.

Proposition 2. The intrinsic priors for problem (1), derived from \( \pi^N_1 \) and \( \pi^{NR}_2 \) (reference prior) are,

\[
\pi^{IR}_1(\mu, \sigma) = \sigma^{-1}, \quad \pi^{IR}_2(\mu, \sigma, \sigma_a) = \pi^{NR}_2(\mu, \sigma, \sigma_a).
\]

That is, in this case the intrinsic and reference prior coincide.

Proof. If \( n = 1 \), then \( g^R_n \) is a proper distribution and then a minimal training sample corresponds to a sample \( y^* \) with \( n = 1, k = 2 \). Moreover, in this case, the associated Bayes factor \( B^N_{12}(y^*) = 1 \) and then, by definition (see (12)), \( \pi^{IR}_2 = \pi^{NR}_2 \).

Surprisingly, \( g^R_n \) is not a proper distribution (except for the simple case \( n = 1 \)) and hence obviously, the intrinsic priors derived from reference priors (at least for the considered ordered group) cannot be used in the model selection problem (1).

As it is shown in the next result, the intrinsic priors derived from Jeffreys prior, can be expressed in a closed form, using the hypergeometric function (Abramowitz and Stegun, 1970):

\[
F(a, b, c, z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 s^{b-1}(1-s)^{c-b-1}(1-s z)^{-a} ds.
\]
Proposition 3. The intrinsic priors for problem (1), derived from \( \pi_1^N \) and \( \pi_2^{NJ} \) (Jeffreys prior) are,

\[
\pi_1^{IJ}(\mu, \sigma) = \sigma^{-1}, \quad \pi_2^{IJ}(\mu, \sigma, \sigma_a) = \sigma^{-1} \pi_1^{IJ}(\sigma_a | \mu, \sigma),
\]

(16)

where

\[
\pi_1^{IJ}(\sigma_a | \mu, \sigma) = \frac{1}{\sigma} f_n^{IJ}(\frac{\sigma_a}{\sigma}), \quad f_n^{IJ}(t) = \frac{16}{3 nt^2 + 1} F\left(\frac{1}{2}, 1, \frac{5}{2}, -2t^2\right).
\]

(17)

Proof. See the Appendix. \( \square \)

The first thing to remark is that \( f_n^{IJ} \) is not a proper density. Despite the fact this function is integrable in \( \mathcal{R}^+ \) (\( f_n^{IJ}(t) = O(t^{-2}) \) as \( t \) tends to infinity), the quantity \( b^{IJ}(n) = \int_0^\infty f_n^{IJ}(t) dt \) is not one. This implies that the intrinsic prior produces a biased Bayes factor. Because \( b^{IJ}(n) \leq b^{IJ}(1) \approx 6.6 \) and \( b^{IJ}(n) \) decreases with \( n \), the bias is positive in favor of \( M_2 \) (\( b^{IJ} > 1 \)) for small values of \( n \) (\( n \leq 12 \) to be more precise) while the bias is positive in favor of \( M_1 \) (\( b^{IJ} < 1 \)) for \( n \) moderate or large (\( n \geq 13 \)). This behavior clearly precludes the direct use of the intrinsic (Jeffreys based) priors in this problem. That the intrinsic priors sometimes lead to biased procedures was pointed out by Berger and Pericchi (1996b). The bias in intrinsic priors has been identified with the use of the Jeffreys prior as the noninformative prior. Furthermore, as Berger and Pericchi (1996b) show, in the normal linear models, the intrinsic prior obtained using the Jeffreys prior is not a proper density while that obtained from the reference prior is a proper density. As the authors claim: “an adjustment is needed for the Jeffreys prior case.”

The (adjusted) intrinsic prior for \( M_2 \) for problem (1) is of the form in (16) but now:

\[
f_n^{I^*}(t) = c^I(n) \frac{t}{nt^2 + 1} F\left(\frac{1}{2}, 1, \frac{5}{2}, -2t^2\right),
\]

(18)

where \( c^I(n) = \left(\int_0^\infty \frac{t}{1 + nt^2} dt\right) F\left(1/2, 1, 5/2, -2t^2\right) \). We have represented \( f_n^{I^*} \), for the case \( n = 5 \), in Figure 3.

4 Examples

We have analyzed the proposals in the previous section for three popular data sets, studied in the context of random effects models. We provide a very brief description of these data sets (details can be found in the references).

- Dyestuff data (Box and Tiao, 1973). The yield of dyestuff in grams of standard color of \( n = 5 \) samples from each of \( k = 6 \) batches of raw material is measured. In this data set, \( SSA = 56357.5 \) and \( SST = 115187.5 \).

- Box and Tiao simulated data (Box and Tiao, 1973). This data set, with \( n = 5 \) and \( k = 6 \), was simulated so that the between mean square \( SSA/(k - 1) \) is less than the within mean square \( SSE/(k(n - 1)) \) and hence, the usual frequentist method for estimation of variances would yield a negative estimated variance. In this data set, \( SSA = 41.68 \) and \( SST = 400.38 \).
Table 1: Analyzed data sets; Bayes factors (5) for different proposals of \( f_n \) (second to fifth rows); median intrinsic Bayes factor (sixth row) and calibrated p-Value (seventh row). Key (according to the classification of Kass and Raftery, 1995): ‘Pi’ means positive evidence against \( M_i \); ‘Si’ means strong evidence against \( M_i \); ‘VSi’ means very strong evidence against \( M_i \).

- Strength of a fabric (Montgomery, 1991). In a textile company, \( k = 4 \) looms are selected random, and \( n = 4 \) strengths determinations are made on the fabric manufactured on each loom. In this data set, \( SSA = 89.19 \) and \( SST = 111.94 \).

For these three data sets, we have computed the Bayes factor \( B_{21} \) in (5) for the proposals: i) \( f_n = f^{WG} \) in (8) (we denote \( B_{21}^{WG} \) the associated Bayes factor); ii) \( f_n = f_n^D \) in (11) for the cases \( q = 0.75 \) (\( B_{21}^{D.0.75} \)), \( q = 1 \) (\( B_{21}^{D.1} \)) and \( q = 1.25 \) (\( B_{21}^{D.1.25} \)); and iii) \( f_n = f_n^{I*} \) in (18) (\( B_{21}^{I*} \)). We also have calculated the median intrinsic Bayes factor (\( MIBF_{21} \)). We used \( MIBF \) (instead of say the arithmetic version) because for these data sets with small amount of data, the arithmetic intrinsic Bayes factor behaved very strangely. The median version of the intrinsic Bayes factor has shown to work (see Berger and Pericchi, 1998) well in a variety of difficult situations (like small sample sizes). Finally, we have computed, for each data set, a calibration of the p-Value (\( p_v \)) suggested by Sellke, Bayarri and Berger (2001), namely

\[
\overline{B}_{21} = \begin{cases} 
1/(-e p_v \log(p_v)), & p_v < e^{-1}, \\
1, & \text{otherwise}.
\end{cases}
\]

This calibration is inspired by the fact that the p-Value is far from being a measure of evidence in favor of a hypothesis (Berger and Delampady, 1987; Berger and Sellke, 1987), but that such simple transformation \( \overline{B}_{21} \) can be viewed as an objective upper bound on the posterior odds of \( M_2 \) to \( M_1 \). Table 1 contains all these computations. We have also included the classification of evidence provided by the Bayes factor proposed by Kass and Raftery (1995) (similar to the one proposed by Jeffreys, 1961).

The main conclusion is that the different Bayes factors provide similar results leading to virtually the same classification of the evidence (in a qualitative sense). Nevertheless, there are some subtle interesting differences among the results. The Bayes factor obtained with the DB prior with \( q = 1 \), \( B_{21}^{D.1} \), is very close to \( B_{21}^{I*} \), the Bayes factor obtained with the (modified) intrinsic prior. \( B_{21}^{D.0.75} \).
and $MIBF_{21}$ (more pronounced in this case) exhibit less evidence against the simple model $M_1$ (are “more conservative” using classic terminology). On the opposite, $B_{21}^{D,1.25}$ and $B_{21}^{WG}$ provide higher levels of evidence against the simple model ($B_{21}^{WG}$ being close to $B_2$, so perhaps providing too much evidence in favor of the more complex model $M_2$). Note that $B_{21}^{D,1}$ and $B_{21}^{I\ast}$ seem to be in the centre of these extremes. Finally, the values of $B_{21}^D$ for the three values of $q$ used were reasonably close to each other. Nevertheless, the results reinforce the (Jeffreys-Zellner-Siow based) proposal of $q = q^\ast + 1/2$ as a sensible automatic election for the exponent in the definition of divergence based priors.

5 The unbalanced case

The unbalanced random effects model is defined as

$$M : Y_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, k, j = 1, \ldots, n_i,$$

where $e_{ij} \sim N(0, \sigma^2)$, iid for all $i, j$ and $a_i \sim N(0, \sigma_a^2)$ iid for all $i$. The models under competition are the same as in (1).

Using similar arguments to those in Section 2, it seems fine to use

$$\pi_1(\mu, \sigma) = \sigma^{-1}$$

and

$$\pi_2(\mu, \sigma, \sigma_a) = \sigma^{-1} \pi(\sigma_a | \mu, \sigma), \quad \pi(\sigma_a | \mu, \sigma) = \sigma^{-1} f_n^u(\sigma_a/\sigma), \quad (19)$$

where $f_n^u$ is a proper function, possibly dependent on $n = (n_1, \ldots, n_k)$. In this situation, the Bayes factor can still be expressed using a univariate integral:

$$B_{21} = N^{1/2} \int \prod_{i=1}^k (1 + n_i t^2)^{-1/2} \left( \sum_{i=1}^k \frac{n_i}{1 + n_i t^2} \right)^{-1/2} \left( \frac{S(t)}{SST} \right)^{(N-1)/2} f_n^u(t) dt,$$

where $N = \sum_{i=1}^k n_i$,

$$S(t) = \sum_{i=1}^k \left( \sum_{j=1}^{n_i} y_{ij}^2 - \frac{t^2}{1 + n_i t^2} \frac{n_i}{y_t} \right)^2 - \left( \sum_{i=1}^k \frac{n_i y_t}{1 + n_i t^2} \right)^2,$$

and

$$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2, \quad \bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}, \quad \bar{y} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}.$$

Now, the exact form of $f_n^u$ has to be determined. One can still argue that the same proposal of $f_n^u = f$ (not dependent on $n$, like the one by Westfall and Gönem, 1996), which works fine for the balanced case, will work in the unbalanced situation. Nevertheless, it seems difficult to justify a prior that does not take into account the design of the experiment (a very valuable prior information). Besides, automatic known priors, like the intrinsic priors are very difficult (if possible) to determine in this situation. On the contrary, one of the attractiveness of the divergence based (DB) priors is that they are generated by a general simple expression. In fact, there is no difficulty
in finding the expression for the DB prior in the unbalanced situation. It is very easy to obtain that the DB prior for the new parameter in the unbalanced case is

$$
\pi^{D,u}(\sigma_a | \mu, \sigma) = c(n, q)\sigma^{-1}\left(1 + \frac{1}{2} \sum_{i=1}^{k} n_i \frac{n_i}{N}\left( \frac{\sigma}{\sigma_a} \right)^2 \left( \frac{\sigma}{\sigma_a} \right)^2 + n_i \right)^{-q}, \quad q > q^*,
$$

where $q^* = 1/2$ and

$$
c(n, q)^{-1} = \int_0^\infty \left(1 + \frac{1}{2} \sum_{i=1}^{k} n_i \frac{n_i s^4}{N(1 + n_i s^2)} \right)^{-q} ds.
$$

Note that this prior is of the form in (19) with

$$
f_n^{D,u}(t) = c(n, q)\left(1 + \frac{1}{2} \sum_{i=1}^{k} n_i \frac{n_i t^4}{N(1 + n_i t^2)} \right)^{-q}, \quad q > q^*.
$$

Again, we recommend $q = q^* + 1/2$ as sensible election for $q$.

6 Summary and conclusions

The choice of objective prior distributions for model selection, in the context of random effects models, has been carefully studied. The same noninformative prior has been proposed, under both models, for the common parameters $\mu, \sigma$, namely $\pi^N(\mu, \sigma) = \sigma^{-1}$. Note that $\pi^N$ is the reference prior under $M_1$. Conditionally on $\mu$ and $\sigma$, the new parameter $\sigma_a$, is assumed to have under $M_2$ a proper prior (to be determined) scaled by $\sigma$. A number of important arguments have been considered, justifying the sensibleness of prior distributions with the requirements above. The determination of the prior for $\sigma_a$ motivated the rest of the paper.

As the main contribution of the paper, we have derived two different kinds of priors for the new parameter: the intrinsic prior (Berger and Pericchi, 1996a) and the divergence based (DB) prior. The DB prior belongs to a recently introduced class of priors (see García-Donato, 2003), dependent on a parameter $q$. The Bayes factor associated with the intrinsic prior is biased, so a correction was introduced in order to avoid this bias. With respect to the DB prior, we were able to show that the associated Bayes factor is consistent (asymptotically selects the true model) when $k \to \infty$ or $n \to \infty$, independently of the value of $q$. Westfall and Gönen (1996) (WG) proposed a prior for the new parameter based on an interesting (but rather ad hoc) reasoning. Their prior was also considered in the examples analyzed.

The different priors have been illustrated on three data sets. The resulting Bayes factors are compared with other proposals (like the median intrinsic Bayes factor, MIBF). For comparison purposes, we calculated also $\bar{B}_{21}$, a measure based on the $p$-Value proposed by Sellke, Bayarri and Berger (2001), which can be interpreted as an upper bound on the evidence in favor of $M_2$. The primary conclusion is that the results, for the analyzed priors, do not change substantially. Nevertheless, some subtle interesting differences between the proposals arise. The (modified) intrinsic prior and the DB prior (for $q = 1$, our preferred automatic election for $q$) provided similar results.
Finally, we have analyzed the unbalanced scenario (one in which, for instance, the expressions for the intrinsic priors are very involved and hence very difficult to obtain) to explicitly demonstrate the simplicity of the methodology of the DB priors. Expressions for the unbalanced version of the DB priors are derived as well as those for the corresponding Bayes factor.

7 Acknowledgements

The work of García-Donato was supported in part by the Spanish Ministry of Science and Education, under grant MTM2004-03290.

Appendix. Proofs

Proof of Proposition 1. It is straightforward to show that

\[ m_1(y) = \int p_1(y \mid \mu, \sigma)\sigma^{-1}d\mu d\sigma = \frac{(SST\pi)^{-nk-1/2}}{2}(nk)^{-1/2}\Gamma\left(\frac{nk-1}{2}\right). \]

Besides,

\[ m_2(y) = \int p_2(y \mid \sigma, \sigma_a)\sigma^{-1}\sigma^{-1}f_n\left(\frac{\sigma_a}{\sigma}\right)d\sigma d\sigma_a, \]

where

\[ p_2(y \mid \sigma, \sigma_a) = \int p_2(y \mid \mu, \sigma, \sigma_a)d\mu = (2\pi)^{-nk-1/2}(nk)^{-1/2}\left(\sigma^2 + n\sigma_a^2\right)^{-k-1/2}\sigma^{-(n-1)k} \exp\left\{-\frac{1}{2\sigma^2}\left(SST - \frac{n\sigma_a^2 SSA}{\sigma^2 + n\sigma_a^2}\right)\right\}. \]

Now, apply the change \( \sigma = \sigma, t = \sigma_a/\sigma \), with Jacobian \( \sigma \), in the integral (20) to obtain

\[ m_2(y) = (2\pi)^{-nk-1/2}(nk)^{-1/2}\int \sigma^{-nk}\exp\left\{-\frac{1}{2\sigma^2}\left(SST - \frac{nt^2 SSA}{1 + nt^2}\right)\right\}(1 + nt^2)^{-k-1/2}f_n(t)d\sigma dt, \]

from which the result follows easily.

Proof of Theorem 1. We prove a more general result.

Proposition 4. For any proper density \( f_n(t) = c_n k_n(t) \), where \( c_n \) does not depend on \( t \), satisfying

\[
\begin{align*}
\text{C1} & \quad k_n(\sqrt{t/n}) \leq A(t) \text{ for all } t \geq 0, \text{ where } A \text{ is bounded in } [0, \infty) \text{ and } A(t) = O(t^{-r}), \text{ with } r > 1/2; \\
\text{C2} & \quad \lim_{n \to \infty} k_n(\sqrt{t/n}) > 0 \text{ for all } t \geq 0; \\
\text{C3} & \quad \lim_{n \to \infty} c_n = c_* > 0.
\end{align*}
\]

the Bayes factor \( B_{21} \) in (5) is consistent under Scenario I and Scenario II.
It is very easy to verify that Conditionas C1 and C2 hold for $f^D$ in (11). That $f^D$ also satisfies C3 is proved in Lemma 1, completing the proof of consistency of DB prior.

**Lemma 1.** Let $c(n,q)$ be defined as in (10), with $q > 1/2$, then

$$
\lim_{n \to \infty} c(n,q) = \frac{\sqrt{\pi}}{\sqrt{2}} \frac{\Gamma(q-1/2)}{\Gamma(q)}
$$

(21)

**Proof.** Note

$$
\left(1 + \frac{1}{2} \frac{nt^2}{n + t^{-2}}\right)^{-q} \leq \left(1 + \frac{1}{2} \frac{t^2}{1 + t^{-2}}\right)^{-q},
$$

(22)

and the quantity on the right is integrable because $q > 1/2$. Now apply the dominated convergence theorem (DCT) and the result holds.

**Lemma 2.** Under $M_1$, for any fixed $k > 1$, the quantity

$$
\left(1 - \frac{SSA}{SSE}\right)^{-(nk-1)/2}
$$

is a.s. bounded as $n \to \infty$.

**Proof.** Let $A_1 = \{\omega \in \Omega : \frac{SSE}{n} \to k - 1 \text{ as } n \to \infty\}$. By the strong law of large numbers (SLLN), $Pr(A_1) = 1$ under $M_1$. Besides, the distribution of SSA under $M_1$ is free of $n$.

Using simple limit arguments and for any $\omega \in A_1$,

$$
\left(1 - \frac{SSA}{SSE + SSA}\right)^{-(nk-1)/2} \to \exp\left\{ \frac{k SSA}{2(k-1)} \right\},
$$

which is a.s. bounded.

**Lemma 3.** let $k_n$ be any function satisfying Condition C2. For any fixed $k > 1$, the integral

$$
\int_0^\infty (1 + t)^{-\frac{k-1}{2}} t^{-1/2} k_n\left(\frac{\sqrt{t}}{\sqrt{n}}\right) dt,
$$

is bounded as $n \to \infty$.

**Proof.** Clearly

$$
(1 + t)^{-\frac{k-1}{2}} t^{-1/2} k_n\left(\frac{\sqrt{t}}{\sqrt{n}}\right) \leq (1 + t)^{-\frac{k-1}{2}} t^{-1/2} A(t) = h^*(t), \text{ say.}
$$

It is easy to show that under condition C1, $h^*$ is an integrable function. By the DCT:

$$
\lim_{n \to \infty} \int_0^\infty (1 + t)^{-\frac{k-1}{2}} t^{-1/2} k_n\left(\frac{\sqrt{t}}{\sqrt{n}}\right) dt = \int_0^\infty \lim_{n \to \infty} (1 + t)^{-\frac{k-1}{2}} t^{-1/2} k_n\left(\frac{\sqrt{t}}{\sqrt{n}}\right) dt \leq \int_0^\infty h^*(t) dt,
$$

completing the proof.
Lemma 4. Under $M_2$, for any $u > 0$ and fixed $k > 1$, the quantity

$$n^{-1/2}(1 - \frac{u}{1+u} \frac{SSA}{SST})^{-\frac{k-1}{2}} \to \infty,$$

a.s. as $n \to \infty$.

Proof. It suffices to show that, as $n \to \infty$,

$$(1 - \frac{u}{1+u} \frac{SSA}{SST})^{-1} \to H,$$

where $H$ is a.s. greater than 1. Let

$$A_2 = \{\omega \in \Omega : \frac{SSE}{\sigma^2 + n\sigma_a^2} \to \frac{k-1}{\sigma_a^2}, \text{ as } n \to \infty\}.$$

Clearly, by the SLLN $Pr(A_2) = 1$, under $M_2$. Write

$$(1 - \frac{u}{1+u} \frac{SSA}{SST})^{-1} = (1 + u)^{-1}(1 + u \frac{SSA}{\sigma^2 + n\sigma_a^2 + V}),$$

where $V = SSA/(\sigma^2 + n\sigma_a^2)$ follows, under $M_2$, a chi-squared distribution with $k - 1$ degrees of freedom and hence a distribution free of $n$. Now, for any $\omega \in A_2$,

$$(1 - \frac{u}{1+u} \frac{SSA}{SST})^{-1} \to H = (1 + u)^{-1}(1 + u \frac{k-1}{k-1 + V\sigma_a^2})^{-1}.$$

Because $V\sigma_a^2 > 0$, a.s., then $H > 1$ a.s., completing the proof. \hfill \square

We are now in conditions to prove the Theorem 1. Consider first:

**Scenario I.** The Bayes factor $B_{21}$ can be written as

$$B_{21} = \frac{c_n}{2\sqrt{n}} \int_0^\infty (1 - \frac{u}{1+u} \frac{SSA}{SST}) h_n(u) du,$$

where

$$h_n(u) = (1 + u)^{-\frac{k-1}{2}} u^{-1/2} k_n(\frac{\sqrt{u}}{\sqrt{n}}).$$

**Part I.1.** We show that, under $M_1$, $B_{21} \to 0$ a.s. as $n \to \infty$.

By Condition C3, $\frac{c_n}{2\sqrt{n}}$, tends to 0 as $n \to \infty$, then it suffices to show that

$$\int_0^\infty (1 - \frac{u}{1+u} \frac{SSA}{SST}) h_n(u) du < \infty,$$

a.s. as $n \to \infty$. Note that,

$$\int_0^\infty (1 - \frac{u}{1+u} \frac{SSA}{SST}) h_n(u) du \leq (1 - \frac{SSA}{SST})^{-\frac{k-1}{2}} \int_0^\infty h_n(u) du.$$

The quantity on the right is, according to Lemma 2 and Lemma 3, an a.s. bounded quantity as $n \to \infty$, leading to the proof of the result.
Part I.2. We show that, under $M_2$, $B_{21} \to \infty$ a.s. as $n \to \infty$.

Clearly, for any $M > 0$,

$$\lim_{n \to \infty} B_{21} \geq \lim_{n \to \infty} \int_M^{M+1} n^{-1/2} \left(1 - \frac{u \ SSA}{1 + u \ SST}\right)^{-\frac{nk-1}{2}} c_n h_n(u) \ du$$

$$= \int_M^{M+1} \lim_{n \to \infty} n^{-1/2} \left(1 - \frac{u \ SSA}{1 + u \ SST}\right)^{-\frac{nk-1}{2}} c_n h_n(u) \ du.$$

Because (see Lemma 4)

$$n^{-1/2} \left(1 - \frac{u \ SSA}{1 + u \ SST}\right)^{-\frac{nk-1}{2}} \to \infty,$$

a.s. as $n \to \infty$, it is enough to prove that $\lim_{n \to \infty} c_n h_n(u) > 0$ for any $u > 0$. But,

$$\lim_{n \to \infty} c_n h_n(u) = c^* (1 + u)^{-\frac{k-1}{2}} u^{-1/2} \lim_{n \to \infty} k_n(\sqrt{t/n}),$$

which, under Conditions C2 and C3 is a positive quantity.

Scenario II

Let us denote

$$F_0 = \frac{SSA}{SSE} \frac{k(n-1)}{k-1}.$$

It is easy to show that, by the SLLN, and as $k \to \infty$:

under $M_1$, $F_0 \to 1$ a.s. and under $M_2$, $F_0 \to 1 + n\sigma^2 > 1$ a.s.

(23)

Now, it is known (Westfall and Gönen, 1996) that for any Bayes factor of the form in (5) with $f_n$ independent of $k$, if $F_0 \leq 1$ then $B_{21} \to 0$ as $k \to \infty$ and if $F_0 > 1$ then $B_{21} \to \infty$ as $k \to \infty$. With this in mind plus (23) it can be easily deduced the consistency of (5) under scenario II for $f_n = f^D$ because $f^D$ is independent of $k$.

Proof of Proposition 3. It is easy to note that, with the priors $\pi_1^N$ and $\pi_2^{NJ}$, a minimal training sample corresponds to a sample with $n = 2$ and $k = 2$. Let $y^*$ be a sample from model $p_2$ in (1) with $k = n = 2$, we prove that

$$E_{y^*}^{M_2} (B_{12}^N (y^*) \mid \mu, \sigma, \sigma_a) = \frac{16}{3} F\left(\frac{1}{2}, 1, \frac{5}{2}, -2\sigma^2_a / \sigma^2\right),$$

from which, the Proposition holds easily.

We obtain first $B_{21}^N (y^*)$. Note that the noninformative priors $\pi_1^N$ and $\pi_2^{NJ}$ in (13) are of the same form as in (3) with $f_n(t) = g_n^J(t) = t/(1 + nt^2)$. Despite $g_n^J$ is improper, the expression of Bayes factor in 5 holds and still can be used to obtain (recall $k = n = 2$)

$$B_{21}^N (y^*) = \int_0^{\infty} (1 + 2t^2)^{-1/2} (1 - \frac{SSA}{1 + 2t^2 \ SST})^{-3/2} \frac{t}{1 + 2t^2} \ dt$$

Change: $2t^2 = u$

$$= \frac{1}{4} \int_0^{\infty} (1 + u(1 - \frac{SSA}{SST}))^{-3/2} \ du = \left(2(1 - \frac{SSA}{SST})\right)^{-1}.$$
Now, the quotient $SSA/SST$ can be conveniently expressed as

$$\frac{SSA}{SST} = \frac{F_1(\rho^2 + 1/2)}{1 + F_1(\rho^2 + 1/2)},$$

where $\rho = \sigma_a/\sigma$ and $F_1 = \frac{SSA}{SSE} \frac{2\sigma^2}{2\sigma_a^2 + \sigma^2}$,

and then, as a function of $F_1$ and $\rho$,

$$B_{12}^N(F_1, \rho) = \frac{2}{1 + F_1(\rho^2 + 1/2)}.$$

Under $M_2$ the statistic $F_1 \sim F_{1,2}$, an F-distribution with degrees of freedom 1 and 2, whose density is denoted by $f_{1,2}$. Then,

$$E_{y^*}^{M_2}(B_{12}^N(y^*) | \mu, \sigma, \sigma_a) = \int_0^\infty \frac{2}{1 + f(\rho^2 + 1/2)} f_{1,2}(t) dt$$

$$= \frac{2\sqrt{2}}{\sqrt{2}} \int_0^\infty \frac{1}{1 + t(\rho^2 + 1/2)} \frac{1}{t^{1/2}(1 + t/2)^{3/2}} dt$$

$$= \frac{4}{\sqrt{2}} (\rho^2 + 1/2)^{-1/2} \int_0^1 u^{-1/2} (1 - u)(1 - \frac{\rho^2}{\rho^2 + 1/2} u)^{-3/2} du$$

$$= \frac{4}{\sqrt{2}} (\rho^2 + 1/2)^{-1/2} \frac{\Gamma(1/2)}{\Gamma(5/2)} \frac{\Gamma(3/2)}{\Gamma(5/2)} \frac{\Gamma(1/2)}{\Gamma(5/2)} F\left(\frac{3}{2}, \frac{1}{2}, \frac{5}{2}, \frac{\rho^2}{\rho^2 + 1/2}\right)$$

$$= \frac{16}{3} F\left(\frac{1}{2}, 1, \frac{5}{2}, -2\rho^2\right).$$

Here the third equality holds from changing variable $u = t(\rho^2 + 1/2)/[1 + t(\rho^2 + 1/2)]$. To obtain the last expression above, we have applied known identities of the hypergeometric function (see Abramowitz and Stegun, 1970), completing the proof.

References


