Ann Inst Stat Math (2008) 60:483–498
DOI 10.1007/s10463-007-0117-5

Some remarks on Bayesian inference for one-way ANOVA models

Fabrizio Solari · Brunero Liseo · Dongchu Sun

Received: 14 July 2005 / Revised: 13 November 2006 / Published online: 11 May 2007

Abstract We consider the standard one-way ANOVA model; it is well-known that classical statistical procedures are based on a scalar non-centrality parameter. In this paper we explore both marginal likelihood and integrated likelihood functions for this parameter and we show that they exactly lead to the same answer. On the other hand, we prove that a fully Bayesian testing procedure may provide different conclusions, depending on what is considered to be the real quantity of interest in the model or, said differently, which are the competing hypotheses. We illustrate these issues via a real data example.

Keywords Integrated likelihood · Marginal likelihood · Model choice · Objective Bayes factor · Reference prior

1 Introduction

Analysis of variance (ANOVA) is an extremely important method in exploratory and confirmatory data analysis (Gelman, 2005). Here we focus on the one-way ANOVA, where we assume that data are observed according to the (full) model M_F

F. Solari
ISTAT, Via Magenta 4, Rome 00184, Italy

B. Liseo (✉)
Dip. di studi geoeconomici, linguistici, statistici e storici,
Università di Roma Sapienza, viale del Castro Laurenziano 9, Rome 00161, Italy
e-mail: brunero.liseo@uniromal.it

D. Sun
Department of Statistics, University of Missouri,
146 Middlebush Hall, Columbia, MO 65211-6100, USA

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad j = 1, \ldots, n_i, \;\; i = 1, \ldots, k, \tag{1.1}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ and σ^2 are unknown parameters and ϵ_{ij} are i.i.d. $N(0, \sigma^2)$. Classical analysis of variance focuses on the hypotheses test

$$M_0 : \mu_1 = \cdots = \mu_k \;\text{ versus }\; M_1 : \mu_i \neq \mu_j \quad \text{for at least one pair } (i, j) \tag{1.2}$$

and it is usually based on the statistic

$$F = \frac{\sum_{i=1}^{k} n_i (\bar{Y}_i - \bar{Y})^2 / (k-1)}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 / (n-k)},$$

where $n = \sum_{i=1}^{k} n_i$, $\bar{Y}_i = n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$ and $\bar{Y} = n^{-1} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij}$.

It is well-known that under the null hypothesis M_0, the F statistic follows a $F_{k-1,n-k}$ distribution. Under the alternative hypothesis M_1, the F statistic is distributed according to a non-central $F_{k-1,n-k,\lambda}$ distribution, whose non-centrality parameter λ is given by

$$\lambda = \frac{1}{\sigma^2} \sum_{i=1}^{k} n_i (\mu_i - \bar{\mu})^2,$$

with $\bar{\mu} = n^{-1} \sum_{i=1}^{k} n_i \mu_i$.

The above testing procedure is a clear and classical example of how the problem of eliminating nuisance parameters is handled in frequentist statistics. Although the full model M_F has $k+1$ parameters, a scalar test statistic is constructed to compare different values of the scalar non-centrality parameter λ. In a certain sense, the classical test acts as if we would have observed *only* the marginal experiment, which provides the quantity F, discarding all the information about the single mean treatments, $(\bar{y}_1, \ldots, \bar{y}_k)$.

It is then apparent that there is a loss of information in this procedure. However, it is not clear *how much* information is lost, and this paper is an attempt to clarify these issues. Bertolino et al. (1990) have shown that the classical test can be reinterpreted in terms of a marginal likelihood of the parameter $\tau = \lambda / n$; we will briefly recall their approach in Sect. 2. In Sect. 3 we perform an integrated likelihood approach to the problem and show that, using a conditional reference prior (Berger and Bernardo 1992), the resulting integrated likelihood is exactly equal to the marginal likelihood of Bertolino et al. (1990). This equivalence amounts to say, in our opinion, that when τ is the real quantity of interest, the classical test, or at least its marginal likelihood counterpart, is a correct report of the information available. In Sect. 4 we derive the class of first-order matching priors for the parameter of interest and perform some comparisons among different priors in terms of frequentist coverage. In Sect. 5 we also perform an objective Bayes analysis for comparing different models. We show that the classical F test (and its likelihood counterpart) correspond to a specific model

comparison that may be or may be not the appropriate test in one-way ANOVA problems; we illustrate our findings via a simple real data example.

2 Marginal likelihood approach

The great simplicity of the standard ANOVA model relies on the transformation of a multiparametric problem into an equivalent scalar one, where the only parameter involved is the non-centrality parameter λ, i.e. rewriting the hypotheses test (1.2) as

$$M_0 : \lambda = 0 \quad \text{versus} \quad M_1 : \lambda > 0. \tag{2.1}$$

Then, frequency-based solutions to the ANOVA problem consist only in the computation of the sample realization F_{obs} of the F statistic. On the other hand, Bayesian solutions are complicated because of the possibly high dimension of the parameter $(\boldsymbol{\mu}, \sigma^2)$. Bertolino et al. (1990) propose to consider F_{obs} as the actual result of a marginal experiment and then to use the sampling distribution of the F statistic as a marginal likelihood.

They also suggest that it is more convenient to work with the parameter

$$\tau = \frac{\lambda}{n}$$

instead of λ. This choice has two advantages. First, τ has a clear interpretation as the ratio of between and within variances; second, it depends on the sample size n only through the relative frequencies n_i/n. Denoting by $d_1 = k - 1$ and $d_2 = n - k$ the degrees of freedom of the F statistic, Bertolino et al. (1990) write the marginal likelihood of τ as

$$L(\tau) \propto \sum_{j=0}^{\infty} p_j\left(\frac{n\tau}{2}\right) \frac{\Gamma\left(\frac{d_1+d_2}{2}+j\right)}{\Gamma\left(\frac{d_1}{2}+j\right)} \left(\frac{d_1 F_{\text{obs}}}{d_2 + d_1 F_{\text{obs}}}\right)^j, \tag{2.2}$$

where, for $j = 1, 2, \ldots$, $p_j(z) = \exp\{-z\} z^j / j!$.

It is useful, for reasons that will be clear later, to rewrite $L(\tau)$ as a function of the Kummer confluent Hypergeometric function $M(a, b, z)$ (Abramowitz and Stegun 1964); from (2.2) it immediately follows that

$$\begin{aligned} L(\tau) &\propto \sum_{j=0}^{\infty} p_j\left(\frac{n\tau}{2}\right) \left(\frac{d_1+d_2}{2}\right)_j \left(\frac{d_2}{2}\right)_j^{-1} \left(\frac{d_1 F_{\text{obs}}}{d_2 + d_1 F_{\text{obs}}}\right)^j \\ &\propto \exp\left\{-\frac{n\tau}{2}\right\} M\left(\frac{d_1+d_2}{2}, \frac{d_2}{2}, \frac{n\tau}{2} \frac{d_1 F_{\text{obs}}}{d_2 + d_1 F_{\text{obs}}}\right), \end{aligned}$$

where $(a)_j = \Gamma(a + j)/\Gamma(a)$ is the Pochhammer's symbol.

Making use of an asymptotic approximation for $L(\tau)$, Solari (2002) shows that a noninformative prior for the parameter τ is given by $\pi(\tau) \propto 1/\sqrt{\tau}$. Bertolino and Racugno (1994) perform a robust Bayesian analysis based on the marginal likelihood $L(\tau)$. Their results seem to be strongly dependent on the choice of the function of the parameter $(\boldsymbol{\mu}, \sigma^2)$ used for reformulating the hypotheses test (1.2) in the form (2.1) (Solari 2002).

3 Integrated likelihood approach

In this section we derive an integrated likelihood function for the parameter τ. To do so, in Sect. 3.1 we first propose a one-to-one reparameterization of $(\boldsymbol{\mu}, \sigma^2)$, namely $(\tau, \boldsymbol{\xi})$, where $\boldsymbol{\xi}$ is a vector of suitable nuisance parameters; second, in Sect. 3.2, we apply the reference prior algorithm. We show that, no matter what is the order of importance of the parameters, the marginal reference prior for the nuisance parameter $\boldsymbol{\xi}$ is always the same, while the marginal prior on τ may change; finally in Sect. 3.3 we obtain the integrated likelihood function

$$\tilde{L}(\tau) \propto \int_{\boldsymbol{\xi}} L(\tau, \boldsymbol{\xi}) \pi(\boldsymbol{\xi}|\tau) \mathrm{d}\boldsymbol{\xi}$$

in a closed form.

3.1 Reparameterization of $(\boldsymbol{\mu}, \sigma^2)$

First we re-express model (1.1) in matrix notation. Let $\mathbf{1}_p$ and $\mathbf{0}_p$ denote, respectively, the $p \times 1$ vectors of 1's and 0's. Thus,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{\epsilon},$$

where the $n \times k$ design matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{n_k} & \mathbf{0}_{n_k} & \cdots & \mathbf{1}_{n_k} \end{bmatrix}$$

and $\boldsymbol{\epsilon}$ is $N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$. Then, the parameter τ can be re-written as

$$\tau = \frac{1}{\sigma^2} (\boldsymbol{\mu} - \bar{\mu}\mathbf{1}_k)^t \boldsymbol{R} (\boldsymbol{\mu} - \bar{\mu}\mathbf{1}_k) = \frac{1}{\sigma^2} \boldsymbol{\mu}^t \boldsymbol{D}^t \boldsymbol{R} \boldsymbol{D} \boldsymbol{\mu},$$

where $\boldsymbol{R} = n^{-1} \operatorname{diag}(n_1, n_2, \ldots, n_k)$ and

$$\boldsymbol{D} = \frac{1}{n}\begin{bmatrix} n-n_1 & -n_2 & \cdots & -n_k \\ -n_1 & n-n_2 & \cdots & -n_k \\ \vdots & \vdots & \vdots & \vdots \\ -n_1 & -n_2 & \cdots & n-n_k \end{bmatrix}.$$

Note that the parameter of interest can be also expressed as $\tau = (n\sigma^2)^{-1} \sum_{i=1}^{k} n_i\mu_i(\mu_i - \bar{\mu})$ or, using matrix notation,

$$\tau = \frac{1}{\sigma^2}\,\boldsymbol{\mu}^t\boldsymbol{R}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1}_k) = \frac{1}{\sigma^2}\,\boldsymbol{\mu}^t\boldsymbol{R}\boldsymbol{D}\boldsymbol{\mu}.$$

Now consider the following steps.

Step 1 Since the k-dimensional vector $(\boldsymbol{\mu} - \bar{\mu}\mathbf{1}_k)$ lies in a $(k-1)$-dimensional subspace of R^k, we need to transform $(\boldsymbol{\mu} - \bar{\mu}\mathbf{1}_k)$ into a $(k-1)$-dimensional vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{k-1})^t$. As $\mathrm{rank}(\boldsymbol{RD}) = \mathrm{rank}(\boldsymbol{D}) = k-1$, there exists a $(k-1) \times k$ matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^t\boldsymbol{P} = \boldsymbol{RD}$. Setting $\boldsymbol{\theta} = \boldsymbol{P}\boldsymbol{\mu}$, we have

$$\boldsymbol{P}^t\boldsymbol{\theta} = \boldsymbol{P}^t\boldsymbol{P}\boldsymbol{\mu} = \boldsymbol{RD}\boldsymbol{\mu} = \boldsymbol{R}(\boldsymbol{\mu} - \bar{\mu}\mathbf{1}_k)$$

and hence

$$\boldsymbol{\mu} = \boldsymbol{R}^{-1}\boldsymbol{P}^t\boldsymbol{\theta} + \bar{\mu}\mathbf{1}_k.$$

The Jacobian matrix $\boldsymbol{H}_1$ of the transformation from $(\boldsymbol{\mu}, \sigma^2)$ to $(\boldsymbol{\theta}, \sigma^2, \bar{\mu})$ is then

$$\boldsymbol{H}_1 = \left[\frac{\partial(\boldsymbol{\mu}, \sigma^2)}{\partial(\boldsymbol{\theta}, \sigma^2, \bar{\mu})}\right] = \begin{bmatrix} \boldsymbol{P}\boldsymbol{R}^{-1} & \mathbf{0}_{k-1} \\ \mathbf{0}_k^t & 1 \\ \mathbf{1}_k^t & 0 \end{bmatrix}.$$

Step 2 Now, switching to polar coordinates as in Berger et al. (1998), we can reparameterize $\boldsymbol{\theta}$ in the new parameters $\eta = \boldsymbol{\theta}^t\boldsymbol{\theta} = n^{-1}\sum_{i=1}^{k} n_i(\mu_i - \bar{\mu})^2$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{k-2}) \in \boldsymbol{\Psi} \equiv (0, \pi)^{k-3} \times (0, 2\pi)$, that is,

$$\begin{aligned} \theta_1 &= \sqrt{\eta}\cos\psi_1, \\ \theta_2 &= \sqrt{\eta}\sin\psi_1\cos\psi_2, \\ &\vdots \\ \theta_{k-2} &= \sqrt{\eta}\sin\psi_1 \cdots \sin\psi_{k-3}\cos\psi_{k-2}, \\ \theta_{k-1} &= \sqrt{\eta}\sin\psi_1 \cdots \sin\psi_{k-3}\sin\psi_{k-2}. \end{aligned}$$

Then, we have

$$\boldsymbol{H}_2 = \left[\frac{\partial(\boldsymbol{\theta}, \sigma^2, \bar{\mu})}{\partial(\eta, \sigma^2, \bar{\mu}, \boldsymbol{\psi})}\right] = \begin{bmatrix} \eta^{-1/2}\,\boldsymbol{b}^t & 0 & 0 \\ \mathbf{0}_{k-1}^t & 1 & 0 \\ \mathbf{0}_{k-1}^t & 0 & 1 \\ \eta^{1/2}\,\boldsymbol{C} & \mathbf{0}_{k-2} & \mathbf{0}_{k-2} \end{bmatrix},$$

where the $(k-1)$-dimensional vector $\boldsymbol{b}$ and the $(k-2)\times(k-1)$ matrix $\boldsymbol{C}$ are given by

$$\boldsymbol{b} = \eta^{1/2}\,\frac{\partial \boldsymbol{\theta}}{\partial \eta} = \frac{1}{2}(\cos\psi_1, \ldots, \sin\psi_1 \ldots \sin\psi_{k-2})^t$$

and

$$\boldsymbol{C} = \eta^{-1/2}\,\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\psi}} = \begin{bmatrix} -\sin\psi_1 & \cdots & \cos\psi_1 \sin\psi_2 \times \cdots \times \sin\psi_{k-2} \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \sin\psi_1 \times \cdots \times \sin\psi_{k-3} \cos\psi_{k-2} \end{bmatrix}.$$

Step 3 Since $\tau = \eta/\sigma^2$, the Jacobian matrix $\boldsymbol{H}_3$ of the transformation from $(\eta, \sigma^2, \bar{\mu}, \boldsymbol{\psi})$ to $(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi})$ is given by

$$\boldsymbol{H}_3 = \left[\frac{\partial(\eta, \sigma^2, \bar{\mu}, \boldsymbol{\psi})}{\partial(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi})}\right] = \begin{bmatrix} \sigma^2 & 0 & 0 & \mathbf{0}_{k-2}^t \\ \tau & 1 & 0 & \mathbf{0}_{k-2}^t \\ 0 & 0 & 1 & \mathbf{0}_{k-2}^t \\ \mathbf{0}_{k-2} & \mathbf{0}_{k-2} & \mathbf{0}_{k-2} & \boldsymbol{I}_{k-2} \end{bmatrix}.$$

We have defined a new set of parameters $(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi})$, in which τ is the parameter of interest and $\boldsymbol{\xi} = (\sigma^2, \bar{\mu}, \boldsymbol{\psi})$ is the nuisance parameter. It must be noticed that the ordering in $\boldsymbol{\xi}$ is just one among the possible ordering we could have chosen. However, since $\bar{\mu}$ and $\boldsymbol{\psi}$ contain information about location and direction of $\boldsymbol{\mu}$ respectively, they do not seem to give direct information about the parameter of interest τ. On the other hand, there is a strong relation between τ and σ^2. Consequently, it seems reasonable to consider the parameter ordering $\{\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi}\}$, as we actually did, or, alternatively, $\{\tau, \sigma^2, \boldsymbol{\psi}, \bar{\mu}\}$. In any case, it will be shown that the reference priors do not depend on the ordering of the components of the nuisance parameter $\boldsymbol{\xi}$.

For the new parameter $(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi})$, the information matrix $\boldsymbol{I}(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi})$ is given by

$$\boldsymbol{I}(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi}) = \boldsymbol{H}_3 \boldsymbol{H}_2 \boldsymbol{H}_1 \boldsymbol{I}(\boldsymbol{\mu}, \sigma^2) \boldsymbol{H}_1^t \boldsymbol{H}_2^t \boldsymbol{H}_3^t,$$

where

$$\boldsymbol{I}(\boldsymbol{\mu},\sigma^2)=\begin{bmatrix}\frac{1}{\sigma^2}\boldsymbol{A}^t\boldsymbol{A} & \mathbf{0}_k\\ \mathbf{0}_k^t & \frac{n}{2(\sigma^2)^2}\end{bmatrix}=n\begin{bmatrix}\frac{1}{\sigma^2}\boldsymbol{R} & \mathbf{0}_k\\ \mathbf{0}_k^t & \frac{1}{2(\sigma^2)^2}\end{bmatrix}.$$

Some algebra yields

$$\boldsymbol{I}(\tau,\sigma^2,\bar{\mu},\boldsymbol{\psi})=n\,\mathrm{diag}\left(\begin{bmatrix}\frac{\boldsymbol{b}^t\boldsymbol{b}}{\tau} & \frac{\boldsymbol{b}^t\boldsymbol{b}}{\sigma^2}\\ \frac{\boldsymbol{b}^t\boldsymbol{b}}{\sigma^2} & \frac{\tau\boldsymbol{b}^t\boldsymbol{b}+1/2}{(\sigma^2)^2}\end{bmatrix},\ \frac{1}{\sigma^2},\ \tau\boldsymbol{C}\boldsymbol{C}^t\right),$$

where $\boldsymbol{C}\boldsymbol{C}^t=\mathrm{diag}(1,\sin^2\psi_1,\ldots,\sin^2\psi_1\cdots\sin^2\psi_{k-3})$. Noting that $\boldsymbol{b}^t\boldsymbol{b}=1/4$, we get

$$\begin{aligned}\boldsymbol{I}(\tau,\sigma^2,\bar{\mu},\boldsymbol{\psi})&=n\,\mathrm{diag}\left(\begin{bmatrix}\frac{1}{4\tau} & \frac{1}{4\sigma^2}\\ \frac{1}{4\sigma^2} & \frac{\tau+2}{4(\sigma^2)^2}\end{bmatrix},\ \frac{1}{\sigma^2},\ \tau\boldsymbol{C}\boldsymbol{C}^t\right)\\ &\equiv n\,\mathrm{diag}\left(\begin{bmatrix}i_{11} & i_{12}\\ i_{21} & i_{22}\end{bmatrix},\ i_{33},\ I_{44}\right).\end{aligned}\tag{3.1}$$

where, the first element of the diagonal in the last formula refers to (τ,σ^2), the second element refers to $\bar{\mu}$ and I_{44} refers to $\boldsymbol{\psi}$.

3.2 Reference priors

3.2.1 Direct approach

Here we derive the one-at-a-time reference prior for $(\tau,\sigma^2,\bar{\mu},\boldsymbol{\psi})$ in the ordering $\{\tau,\sigma^2,\bar{\mu},\boldsymbol{\psi}\}$, in the sense that τ is the parameter of interest and $(\sigma^2,\bar{\mu},\boldsymbol{\psi})$ are the nuisance parameters in descending order of interest. Following the reference prior algorithm (Berger and Bernardo, 1992), the conditional reference priors for $(\tau,\sigma^2,\bar{\mu},\boldsymbol{\psi})$ are given by

$$\begin{aligned}\pi_d(\boldsymbol{\psi}|\tau,\sigma^2,\bar{\mu})&\propto\left|\boldsymbol{I}_{44}\right|^{\frac{1}{2}}\ \propto\ \prod_{i=1}^{k-3}(\sin\psi_i)^{k-i-2},\\ \pi_d(\bar{\mu}|\tau,\sigma^2)&\propto\left|i_{33}\right|^{\frac{1}{2}}\ \propto\ 1,\\ \pi_d(\sigma^2|\tau)&\propto\left|i_{22}\right|^{\frac{1}{2}}\ \propto\ \frac{1}{\sigma^2},\\ \pi_d(\tau)&\propto\left|i_{22}\right|^{-\frac{1}{2}}\left|i_{11}i_{22}-i_{12}i_{21}\right|^{\frac{1}{2}}\ \propto\ \frac{1}{\sqrt{\tau(\tau+2)}}\end{aligned}$$

and it is easy to check that they do not change whatever is the ordering of the nuisance parameters $(\sigma^2,\bar{\mu},\boldsymbol{\psi})$.

In all the above computations, we have considered the vector parameter $\boldsymbol{\psi}$ as a single block. If we split it into its $k-2$ distinct components, the new reference prior for each ψ_i, $i = 1, \ldots, k-2$, has the following interesting form:

$$\pi_d(\psi_i \mid \tau, \sigma^2, \bar{\mu}, \psi_1, \ldots, \psi_{i-1}) \propto \prod_{j=1}^{i-1} \sin \psi_j \propto 1.$$

3.2.2 Reverse approach

Now, we will construct the reverse reference prior, where the reverse ordering of the parameter, namely $\{\boldsymbol{\psi}, \bar{\mu}, \sigma^2, \tau\}$ is considered in the derivation of the prior distributions. This way, the prior distribution of the parameter of interest τ conditionally on the nuisance parameter $(\boldsymbol{\psi}, \bar{\mu}, \sigma^2)$ will be computed first. In this case, we have

$$\begin{aligned} I(\boldsymbol{\psi}, \bar{\mu}, \sigma^2, \tau) &= n \operatorname{diag}\left(I_{44},\ i_{33},\ \begin{bmatrix} i_{11} & i_{12} \\ i_{21} & i_{22} \end{bmatrix}\right) \\ &= n \operatorname{diag}\left(\tau \boldsymbol{C}\boldsymbol{C}^t,\ \frac{1}{\sigma^2},\ \begin{bmatrix} \frac{\tau+2}{4(\sigma^2)^2} & \frac{1}{4\sigma^2} \\ \frac{1}{4\sigma^2} & \frac{1}{4\tau} \end{bmatrix}\right). \end{aligned}$$

It follows that the conditional reference priors are

$$\begin{aligned} &\pi_r(\tau \mid \boldsymbol{\psi}, \bar{\mu}, \sigma^2) \propto \left|i_{11}\right|^{\frac{1}{2}} \propto \frac{1}{\sqrt{\tau}}, \\ &\pi_r(\sigma^2 \mid \boldsymbol{\psi}, \bar{\mu}) \propto \left|i_{22}\right|^{\frac{1}{2}} \propto \frac{1}{\sigma^2}, \\ &\pi_r(\bar{\mu} \mid \boldsymbol{\psi}) \propto \left|i_{33}\right|^{\frac{1}{2}} \propto 1, \\ &\pi_r(\boldsymbol{\psi}) \propto \left|\boldsymbol{I}_{44}\right|^{\frac{1}{2}} \propto \prod_{i=1}^{k-3} (\sin \psi_i)^{k-i-2}. \end{aligned}$$

As before, the final answer is independent of the ordering of the nuisance parameters. Note that the reverse approach leads to the marginal prior $\pi_r(\tau) \propto 1/\sqrt{\tau}$, i.e. the same prior derived in Solari (2002) for the marginal likelihood (2.2).

Again, if one considers $k-2$ groups of parameters $\psi_i, i = 1, \ldots, k-2$, instead of the single group $\boldsymbol{\psi}$, a constant reference prior for the ψ_i's is obtained

$$\pi_r(\psi_i \mid \psi_1, \ldots, \psi_{i-1}) \propto \prod_{j=1}^{i-1} \sin \psi_j \propto 1.$$

3.3 Integrated likelihood

In this section we will compute the integrated likelihood using the reference priors obtained in Sects. 3.2.1 and 3.2.2. It must be noticed that both the direct and the reverse reference prior produce the same prior for the nuisance parameter, i.e.

$$\pi(\sigma^2, \bar{\mu}, \boldsymbol{\psi} \mid \tau) = \pi(\sigma^2, \bar{\mu}, \boldsymbol{\psi}) \propto \frac{1}{\sigma^2} \prod_{i=1}^{k-3} (\sin \psi_i)^{k-i-2},$$

resulting in the same integrated likelihood

$$\tilde{L}(\tau) = \int L(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi})\, \pi(\sigma^2, \bar{\mu}, \boldsymbol{\psi}) \mathrm{d}\sigma^2 \mathrm{d}\bar{\mu}\, \mathrm{d}\boldsymbol{\psi}.$$

Starting from the likelihood function for $(\boldsymbol{\mu}, \sigma^2)$

$$L(\boldsymbol{\mu}, \sigma^2) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu})^t(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu})\right\}$$

and passing through the reparameterization from Step 1 to Step 3, after some algebra we obtain

$$L(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi}) \propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{n\tau}{2} - \frac{n(\bar{\mu} - \bar{y})^2}{2\sigma^2} - \frac{SST}{2\sigma^2} + 2n\left(\frac{\tau}{\sigma^2}\right)^{\frac{1}{2}} \bar{\boldsymbol{y}}^t \boldsymbol{P}^t \boldsymbol{b}\right\},$$

where $SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$ is the total sum of squares and $\bar{\boldsymbol{y}} = (\bar{y}_1, \ldots, \bar{y}_k)^t$ represents the sample mean vector.

Using results in Ferrándiz (1982) to integrate out $\boldsymbol{\psi}$ and standard integration methods for $\bar{\mu}$ and σ^2, we obtain that the integrated likelihood results to be the same as the marginal likelihood (2.2), showing that no loss of information derives from considering only the marginal experiment, at least when τ is the real parameter of interest.

4 Choice among noninformative priors

In the previous section we have shown that no matter what reference prior is used, the likelihood function for the parameter of interest τ turns out to be always the same. Now we discuss *which* reference marginal prior for τ should be used, that is we want to compare the two priors $\pi_d(\tau) = 1/\sqrt{\tau(\tau + 2)}$ and

$\pi_r(\tau) = 1/\sqrt{\tau}$ derived from the two approaches described in the previous sections. Furthermore, a frequentist probability matching prior for τ can be also derived; we illustrate this issue in Sect. 4.1.

4.1 The first-order matching prior for τ

In this section we derive all the first-order matching priors for τ in the sense that one-sided credible intervals and their frequentist coverage probabilities agree up to $O\left(n^{-1}\right)$. For the Fisher information matrix $\boldsymbol{I}$ of $(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi})$ given in (3.1), it is easy to show that

$$\boldsymbol{I}^{-1}(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi}) = \frac{2}{n}\text{diag}\left(\begin{bmatrix} \tau(\tau+2) & -\tau\sigma^2 \\ -\tau\sigma^2 & \sigma^4 \end{bmatrix}, \sigma^2, \frac{1}{\tau}(\boldsymbol{C}\boldsymbol{C}^t)^{-1}\right).$$

Let $\mathbf{e}_1 = (1, 0, 0, \mathbf{0}'_{k-2})'$; then $\mathbf{e}_1'\boldsymbol{I}^{-1}\mathbf{e}_1 = 2\tau(\tau+2)/n$ and

$$\boldsymbol{v} = \frac{\boldsymbol{I}^{-1}\mathbf{e}_1}{\sqrt{\mathbf{e}_1'\boldsymbol{I}^{-1}\mathbf{e}_1}} = \sqrt{\frac{2}{n}}\begin{bmatrix} \sqrt{\tau(\tau+2)} \\ -\sqrt{\frac{\tau}{\tau+2}}\sigma^2 \\ 0 \\ \mathbf{0}_{k-2} \end{bmatrix}.$$

Following Datta and Ghosh (1995), a first-order matching prior $\pi(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi})$ for τ must satisfy the differential equation

$$\frac{\partial}{\partial\tau}\sqrt{\tau(\tau+2)}\,\pi(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi}) - \frac{\partial}{\partial\sigma^2}\sqrt{\frac{\tau}{\tau+2}}\sigma^2\,\pi(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi}) = 0 \qquad (4.1)$$

Let $\pi^* = \sqrt{\tau(\tau+2)}\sigma^2\,\pi(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi})$. Clearly, (4.1) is then equivalent to

$$(\tau+2)\frac{\partial}{\partial\tau}\pi^* - \sigma^2\frac{\partial}{\partial\sigma^2}\pi^* = 0,$$

which has a general solution $\pi^* = g((\tau+2)\sigma^2, \bar{\mu}, \boldsymbol{\psi})$, for some positive and differentiable function $g(\cdot, \cdot, \cdot)$. Then a general solution for (4.1) is

$$\pi(\tau, \sigma^2, \bar{\mu}, \boldsymbol{\psi}) = \frac{1}{\sqrt{\tau(\tau+2)}\sigma^2}g((\tau+2)\sigma^2, \bar{\mu}, \boldsymbol{\psi}).$$

It is easy to see that the reference prior π_d is a matching prior by choosing $g(\cdot) = \prod_{i=1}^{k-3}(\sin\psi_i)^{k-i-2}$, which is a constant on its first two arguments. On the other hand, π_r will not satisfy (4.1) and is not a matching prior.

Table 1 Coverage probability for $\pi_d(\tau \mid F_{\text{obs}})$ and $\pi_r(\tau \mid F_{\text{obs}})$ for $\alpha = 0.05$ for values of τ obtained combining $\boldsymbol{\mu} = (1,2,3)$; $\sigma^2 = 0.25, 0.5, 1$; $n_1 = n_2 = n_3 = 2, 5, 10, 30$

$n = 6, \sigma^2 = 0.25$		$n = 6, \sigma^2 = 0.5$		$n = 6, \sigma^2 = 1$	
π_d	π_r	π_d	π_r	π_d	π_r
0.969	0.954	0.974	0.964	0.996	0.990
$n = 15, \sigma^2 = 0.25$		$n = 15, \sigma^2 = 0.5$		$n = 15, \sigma^2 = 1$	
π_d	π_r	π_d	π_r	π_d	π_r
0.963	0.953	0.958	0.948	0.957	0.949
$n = 30, \sigma^2 = 0.25$		$n = 30, \sigma^2 = 0.5$		$n = 30, \sigma^2 = 1$	
π_d	π_r	π_d	π_r	π_d	π_r
0.954	0.946	0.951	0.944	0.953	0.949
$n = 90, \sigma^2 = 0.25$		$n = 90, \sigma^2 = 0.5$		$n = 90, \sigma^2 = 1$	
π_d	π_r	π_d	π_r	π_d	π_r
0.953	0.946	0.951	0.948	0.954	0.951

4.2 Small sample comparison between $\pi_d(\tau)$ and $\pi_r(\tau)$

To see the small sample frequentist performance of the one-sided $1-\alpha$ credible intervals produced by the use of the two priors, we conduct some simulation study. The two posterior densities can be written as

$$\pi_i(\tau \mid F_{\text{obs}}) = \frac{\pi_i(\tau)\, L(\tau)}{m_i(F_{\text{obs}})}, \quad i = d, r,$$

where

$$m_i(F_{\text{obs}}) = \int_0^{\infty} L(\tau)\pi_i(\tau)\mathrm{d}\tau, \quad i = d, r.$$

While the posterior $\pi_r(\tau \mid F_{\text{obs}})$ can be obtained in closed form, the marginal distribution $m_d(F_{\text{obs}})$ obtained from the use of π_d must be calculated via numerical integration. Resorting to the Gauss hypergeometric series $F(a, b, c, z)$ (see Abramowitz and Stegun 1964), we have

$$m_r(F_{\text{obs}}) = \left(\frac{2\pi}{n}\right)^{\frac{1}{2}} F\left(\frac{d_1 + d_2}{2}, \frac{1}{2}, \frac{d_1}{2}, \frac{d_1 F_{\text{obs}}}{d_2 + d_1 F_{\text{obs}}}\right) < \infty.$$

Since $\pi_d(\tau) < \pi_r(\tau)$ then $m_d(F_{\text{obs}}) < \infty$ and both $\pi_d(\tau \mid F_{\text{obs}})$ and $\pi_r(\tau \mid F_{\text{obs}})$ result to be proper posterior densities.

We now compare the frequentist coverage of the one-sided $1 - \alpha$ credible intervals for τ by Monte Carlo simulation. Tables 1 and 2 illustrate the behavior of $\pi_d(\tau \mid F_{\text{obs}})$ and $\pi_r(\tau \mid F_{\text{obs}})$ when $k = 3$ and $k = 4$, respectively. Parameter values in the simulations were choosing in terms of the original parameter $(\boldsymbol{\mu}, \sigma^2)$.

Table 2 Coverage probability for $\pi_d(\tau \mid F_{\text{obs}})$ and $\pi_r(\tau \mid F_{\text{obs}})$ for $\alpha = 0.05$ for values of τ obtained combining $\boldsymbol{\mu} = (1,2,3,4)$; $\sigma^2 = 0.25, 0.5, 1$; $n_1 = n_2 = n_3 = 2, 5, 10, 30$

$n = 8, \sigma^2 = 0.25$		$n = 8, \sigma^2 = 0.5$		$n = 8, \sigma^2 = 1$	
π_d	π_r	π_d	π_r	π_d	π_r
0.967	0.951	0.962	0.945	0.969	0.959
$n = 20, \sigma^2 = 0.25$		$n = 20, \sigma^2 = 0.5$		$n = 20, \sigma^2 = 1$	
π_d	π_r	π_d	π_r	π_d	π_r
0.964	0.952	0.958	0.948	0.957	0.950
$n = 40, \sigma^2 = 0.25$		$n = 40, \sigma^2 = 0.5$		$n = 40, \sigma^2 = 1$	
π_d	π_r	π_d	π_r	π_d	π_r
0.958	0.948	0.958	0.950	0.956	0.951
$n = 120, \sigma^2 = 0.25$		$n = 120, \sigma^2 = 0.5$		$n = 120, \sigma^2 = 1$	
π_d	π_r	π_d	π_r	π_d	π_r
0.956	0.952	0.956	0.952	0.954	0.950

The simulation study supports the idea that both $\pi_d(\tau \mid F_{\text{obs}})$ and $\pi_r(\tau \mid F_{\text{obs}})$ have good frequentist properties. It must be observed that, for small sample sizes, the coverage is in general higher than the nominal value, while for large values of n, the coverage level is approximately equal to $1 - \alpha$ for both $\pi_d(\tau \mid F_{\text{obs}})$ and $\pi_r(\tau \mid F_{\text{obs}})$. This is not surprising for the direct reference prior, since it has been shown to be a matching prior for τ.

5 Bayesian testing

In this section we illustrate the peculiar behavior of the Bayes factor when both the competing models are "wrong". More precisely we perform two distinct testing procedures, one based on the entire likelihood function $L(\boldsymbol{\mu}, \sigma^2)$ and the other based only on the marginal likelihood $L(\tau)$, and we show that the two analyses can give, in particular situations, quite different results. In both cases we perform objective Bayesian model selection; therefore, we cannot use the usual Bayes factor and we need to adopt some modifications of it, namely the intrinsic Bayes factor (Berger and Pericchi, 1996a,b) and the fractional Bayes factor (O'Hagan, 1995). In the marginal likelihood case, the choice of the training sample sizes must be done with care because the parameter of interest τ actually depends on the n_i/n's; consequently, denoting by n' the subsample size and with n'_i, $i = 1, \ldots, k$, the number of units on each treatment in the subsample, the training sample sizes should be chosen under the following condition

$$\frac{1}{n'\sigma^2} \sum_{i=1}^{k} n'_i(\mu_i - \bar{\mu})^2 = \frac{1}{n\sigma^2} \sum_{i=1}^{k} n_i(\mu_i - \bar{\mu})^2, \tag{5.1}$$

which holds if and only if $n'_i/n' = n_i/n$, $i = 1, \ldots, k$. The rationale behind this choice is that, in this way, τ maintains the same intrinsic meaning both in the

Table 3 Ascorbic acid content expressed in mg

T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9
7.12	4.42	6.49	8.07	8.05	5.09	5.87	6.57	4.13
7.16	5.68	8.09	2.86	5.82	4.57	5.36	5.08	7.31
4.57	5.15	8.79	6.84	2.47	6.06	5.85	5.95	4.47
3.79	3.83	8.44	6.85	3.28	4.87	6.27	7.51	2.53
4.20	3.30	6.11	4.12	5.38	4.52	5.96	3.79	3.96
5.84	4.44	5.17	3.32	3.98	5.08	4.95	4.33	5.30
5.56	3.51	8.13	1.74	6.08	4.29	5.85	3.70	2.66
5.02	4.60	7.58	1.74	6.28	6.19	4.70	5.21	4.12
3.69	4.85	6.47	1.57	5.72	3.45	1.53	4.48	3.54
2.99	4.84	5.45	3.02	2.88	5.85	3.88	5.17	2.98
4.99	5.45	6.18	5.08	6.40	2.51	2.88	4.69	5.08
2.16	4.71	4.34	4.96	4.58	4.93	2.07	2.12	5.15
Mean values								
4.76	4.58	6.77	3.88	5.07	4.78	4.59	4.88	4.27

training sample and in the complete sample. Consequently, the usual solution of considering a subsample of size $k+1$ (with the n_i''s all equal to 1 with the exception of one group taking value 2) is not appropriate here. Things are even more complicated for unbalanced ANOVA models; in this case, in fact, it is not always possible to find subsamples satisfying exactly the relation $n_i'/n' = n_i/n$ for all i's.

We illustrate the main point of this section via the following real data example.

Example 1 (Pompilj and Napolitani, 1954). An experiment is conducted to analyze the possible influence of some types of manuring on the ascorbic acid content in tomatoes. The treatments under study are nine manures obtained as different combinations of calcium nitrate and calcium superphosphate. The data of the experiment are reported in Table 3. Notice that the mean value under the third treatment is sensibly larger than the others.

We are generally interested in the comparison between the two models M_0 : $\mu_1 = \cdots = \mu_k$ and $M_1 : \mu_i \neq \mu_j$ for at least a pair (i, j), or, equivalently,

$$M_0 : Y_{ij} = \mu + \epsilon_{ij}, \quad i = 1, \ldots, 9, \;\; j = 1, \ldots, 12$$

and

$$M_1 = Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \ldots, 9, \quad j = 1, \ldots, 12.$$

We now compare via fractional Bayes factor and via arithmetic, geometric and median intrinsic Bayes factors the two models using (a) the complete likelihood and (b) the marginal likelihood.

The results presented in Table 4 are only apparently contradictory. In fact, the analysis based on the marginal likelihood produces evidence in favor of the

Table 4 Bayes factors relative to M_0 versus M_1 for complete and pseudo-likelihood

Complete likelihood			
$B_{01}^{F}(\mathbf{y})$	$B_{01}^{AI}(\mathbf{y})$	$B_{01}^{GI}(\mathbf{y})$	$B_{01}^{MI}(\mathbf{y})$
2.39905	32.30192	7.67851	8.47163
Marginal likelihood			
$B_{01}^{F}(F_{obs})$	$B_{01}^{AI}(F_{obs})$	$B_{01}^{GI}(F_{obs})$	$B_{01}^{MI}(F_{obs})$
0.03421	0.03814	0.02867	0.03725

In the first case we used the reference prior $\pi_0(\mu,\sigma^2) \propto 1/\sigma^2$ and $\pi_1(\mu_1,\ldots,\mu_9,\sigma^2) \propto 1/\sigma^2$. In the latter case we used, for the larger model, the reference prior $\pi_d(\tau)$

Table 5 Bayes factors relative to M_0 versus M_2 for complete and pseudo-likelihood

Complete likelihood			
$B_{02}^{F}(\mathbf{y})$	$B_{02}^{AI}(\mathbf{y})$	$B_{02}^{GI}(\mathbf{y})$	$B_{02}^{MI}(\mathbf{y})$
0.00023	0.00214	0.00058	0.00068
Marginal likelihood			
$B_{02}^{F}(F_{obs})$	$B_{02}^{AI}(F_{obs})$	$B_{02}^{GI}(F_{obs})$	$B_{02}^{MI}(F_{obs})$
0.00003	0.00004	0.00003	0.00004

The reference priors $\pi_0(\mu,\sigma^2) \propto 1/\sigma^2$ and $\pi_1(\mu_1,\mu_2,\sigma^2) \propto 1/\sigma^2$ and $\pi_d(\tau)$ were used for complete and pseudo-likelihood, respectively

model M_1 while the complete likelihood suggests to choose model M_0. One possible explanation of this phenomenon is that the complete likelihood analysis compares the one-dimensional model M_0 against the k-dimensional model M_1 while in the marginal likelihood analysis the comparison is made between $M_0 : \tau = 0$ against the one-dimensional model $M_1 : \tau > 0$.

It is well-known that Bayesian (and likelihood) inference can be highly misleading when using an incorrect model. The Bayes factor and its modifications are known to select the model that is, in a Kullback–Liebler sense, closer to the true one (Dmochowski, 1996). In the ANOVA set-up when only one out of many treatments is sensibly different from the others, it might happen that the Bayes factor based on the complete likelihood selects the null model simply because it is closer to the true model. On the other hand, this is not the answer we expect when the goal is to check the equivalence of "all" the treatments. In such cases the use of the "marginal" experiment must be preferred.

These considerations suggest, instead, to compare the null model M_0 against

$$M_2 = \begin{cases} Y_{ij} = \mu_1 + \epsilon_{ij}, & i \neq 3, \ j = 1,\ldots,12, \\ Y_{ij} = \mu_2 + \epsilon_{ij}, & i = 3, \ j = 1,\ldots,12. \end{cases}$$

Results are presented in Table 5. Notice that, after eliminating the overparameterization of model M_1, the two analysis produce similar results.

A referee correctly pointed out that our testing scenario is more realistic when the researcher already knows *which* treatment might be different from

the others. We agree on that; however we also believe that the above is a quite common situation and, in any case, when there are no prior *suspects* on which treatment might be different, the test can be used repeatedly. The referee also asked if the same testing scenario can be useful when more than one treatment might be different. This issue may be certainly considered and reformulated in terms of model selection and Bayesian methods are generally particularly useful in this context (George, 2000). This last point reinforces our point of view that the classical ANOVA test (or its marginal likelihood counterpart) should be used as a preliminary test for checking if any of the treatment performs differently. At this stage, in fact, the use of the complete likelihood might be misleading if the alternative model is not well specified. If the null model is rejected by the classical test, then a deeper Bayesian model selection procedure should be performed.

Acknowledgments Dongchu Sun work has been supported by the National Science Foundation grant SES-0351523. Brunero Liseo work has been supported by grant provided by Università di Roma Sapienza. The authors would like to thank the editors and two referees for constructive comments to an early version of the paper. Part of the work was done while Fabrizio Solari was visiting the University of Missouri–Columbia.

References

Abramowitz, M., Stegun, I. A. (1964). *Handbook of Mathematical Functions*. New York:Dover.

Berger, J. O., Bernardo, J. M. (1992). On the development of reference priors. In J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (Eds.), *Bayesian statistics 4*, (pp. 35–60) with discussion. London:Oxford University Press.

Berger, J. O., Liseo, B., Wolpert, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science 14*, 1–28 (with discussion).

Berger, J. O., Pericchi, L. R. (1996a). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association 91*, 109–122.

Berger, J. O., Pericchi, L. R. (1996b). The intrinsic Bayes factor for linear models. In J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (Eds.), *Bayesian statistics 5*, (pp. 23–42). London:Oxford University Press.

Berger, J. O., Philippe, A., Robert, C. P. (1998). Estimation of quadratic functions: noninformative priors for non-centrality parameters. *Statistica Sinica 8*, 359–375.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society B 41*, 113–147 (with discussion).

Bertolino, F., Piccinato, L., Racugno, W. (1990). A marginal likelihood approach to analysis of variance. *The Statistician 39*, 415–424.

Bertolino, F., Racugno, W. (1994). Robust Bayesian analysis of variance and the χ^2–test by using marginal likelihoods. *The Statistician 43*, 191–201.

Datta, G. S., Ghosh, J. K. (1995). On priors providing validity for Bayesian inference. *Biometrika 82*, 37–45.

Dmochowski, J. (1996). Intrinsic priors via Kullback-Liebler geometry. In J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (Eds.), *Bayesian statistics 5*, (pp. 543–549).

Ferrándiz, J. R. (1982). Una solución Bayesiana a la paradoja de Stein. *Trabajos de Estadistica 33*, 31–46.

Gelman, A. (2005). Analysis of variance – Why it is more important than ever. *Annals of Statistics 33*, 1–53 (with discussion).

George, E. (2000). The variable selection problem. *Journal of the American Statistical Association*, *95*, 1304–1308.

Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika 80*, 295–304.
O'Hagan, A. (1995). Fractional Bayes factors for model comparisons. *Journal of the Royal Statistical Society B 57*, 99–138.
Pompilj, G., Napolitani, D. (1954). Piano degli Esperimenti. *La Ricerca Scientifica*, *24*, suppl.
Solari, F. (2002). L'uso di pseudo-verosimiglianze nell'analisi della varianza. Phd thesis, Dipartimento di statistica, probabilità e statistiche applicate, Università di Roma Sapienza.