International Statistical Review

International Statistical Review (2026) doi: 10.1111/insr.70028

On Integral Priors for Multiple Comparison in Bayesian Model Selection

Diego Salmerón[1,2,3], **Juan Antonio Cano**[4] and **Christian P. Robert**[5,6]

[1]*Health and Social Sciences Department, University of Murcia, Murcia, 30120, Spain*
[2]*CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, 28029, Spain*
[3]*Biomedical Research Institute of Murcia Pascual Parrilla-IMIB, Murcia, 30120, Spain*
[4]*Department of Statistics and Operational Research, University of Murcia, Murcia, 30100, Spain*
[5]*Université Paris Dauphine PSL, Paris, France*
[6]*Department of Statistics, University of Warwick, Coventry, CV4 7AL, United Kingdom*
Corresponding to: Diego Salmerón, Health and Social Sciences Department, University of Murcia, Murcia 30120, Spain. Email: dsm@um.es

Summary

Noninformative priors constructed for estimation purposes are usually not appropriate for model selection and testing. The methodology of integral priors was developed to get prior distributions for Bayesian model selection when comparing two models, modifying initial improper reference priors. We propose a generalisation of this methodology to more than two models. Our approach adds an artificial copy of each model under comparison by compactifying the parametric space and creating an ergodic Markov chain across all models that returns the integral priors as marginals of the stationary distribution. Besides the guarantee of their existence and the lack of paradoxes attached to estimation reference priors, an additional advantage of this methodology is that the simulation of this Markov chain is straightforward as it only requires simulations of imaginary training samples for all models and from the corresponding posterior distributions. We present some examples, including situations where other methodologies need specific adjustments or do not produce a satisfactory answer.

Key words: Bayesian model selection; Integral priors; Markov chains; Objective Bayes factor.

1 Introduction

Noninformative (or objective) priors are essential tools for Bayesian analysis, even though the denominations of 'noninformative' and 'objective' are often criticised as being inappropriate (Gelman *et al.*, 2013). The main motivation is in producing a reference prior distribution that impacts as little as possible the resulting posterior distribution (Berger *et al.*, 2024). Further motivations range from closing the inference space as in complete class results (Robert, 2001) to achieving better frequentist properties such as asymptotic coverage (Rousseau *et al.*, 2007).

The main issue with such priors is that there exists an extensive literature on the possible choices and theories of noninformative priors. Following their formalisation by Harold Jeffreys (Jeffreys, 1939), several propositions emerged towards a theory of such priors, with a varying

degree of generality as reported in Kass & Wasserman (1995), Robert (2001), Rissanen (2012), and Berger *et al.* (2024). The most elaborate approach to this day is the concept of reference priors developed by the latter (see also Bernardo, 1979 and Berger & Bernardo, 1992).

These different methods most usually yield an improper prior measure, that is, a measure with infinite mass, which is thus determined up to a positive multiplicative constant since it cannot be normalised into a probability distribution. Most often this feature is not a serious issue for estimation purposes because the constant cancels in the posterior distribution. Unfortunately, in model selection problems such as hypothesis testing where two (or more) models compete for consideration, $M_i: \{f_i(x|\theta_i), \theta_i \in \Theta_i\}, i = 1, 2$, resorting to improper priors $\pi_i^N(\theta_i) = c_i h_i(\theta_i), i = 1, 2$, implies that the optimal Bayesian procedure, namely, the Bayes factor (Jeffreys, 1939)

$$B_{21}^N(x) = \frac{c_2 \int f_2(x|\theta_2) h_2(\theta_2) \mathrm{d}\theta_2}{c_1 \int f_1(x|\theta_1) h_1(\theta_1) \mathrm{d}\theta_1},$$

depends on the arbitrary ratio c_2/c_1. Several general approaches have been proposed for prior construction in Bayesian model selection, among which Bayarri & García-Donato (2008), Bayarri *et al.* (2012), and Fouskakis *et al.* (2015). A review of reference priors for model selection can be found in Consonni *et al.* (2018) as well as in Berger *et al.* (2024).

One of these attempts towards alleviating this issue is the methodology of intrinsic priors (Berger & Pericchi, 1996) that exploit a (minimal) fraction of the data to update the improper priors into proper posteriors. In short, intrinsic priors are defined as the solution $\{\pi_1^I(\theta_1), \pi_2^I(\theta_2)\}$ of a system of two functional equations. Under certain conditions, and assuming that M_1 is nested in M_2, the system of functional equations reduces to a single equation with two functional unknowns, in such a way that once an arbitrary choice is made for $\pi_1^I(\theta_1)$, the dual prior $\pi_2^I(\theta_2)$ is automatically determined. The solution $\{\pi_1^I(\theta_1), \pi_2^I(\theta_2)\}$ proposed by Moreno *et al.* (1998) consists of choosing $\pi_1^I(\theta_1) = \pi_1^N(\theta_1)$, and hence,

$$\pi_2^I(\theta_2) = \pi_2^N(\theta_2) \mathbb{E}_{f_2(z|\theta_2)}(B_{12}^N(Z)) = \int \pi_2^N(\theta_2|z) m_1^N(z) \mathrm{d}z,$$

where z denotes an imaginary training sample, $B_{12}^N(z) = m_1^N(z)/m_2^N(z)$, and $m_i^N(z) = \int f_i(z|\theta_i) \pi_i^N(\theta_i) \mathrm{d}\theta_i, i = 1, 2$. See Pérez & Berger (2002) for more details. However, intrinsic Bayes factors are often not suitable in the non-nested case, see Berger & Mortera (1999) and Moreno (2005). Another proposal consists of using fractional priors, which are again defined as the solution to a system of functional equations (see O'Hagan, 1997 and Moreno, 1997). However, this methodology has received less attention and also faces difficulties (Robert, 2001).

Pérez & Berger (2002) have proposed a novel approach under the name of expected posterior priors distributions. This method can be applied in the general scenario where we have $q \geq 2$ models under consideration. The expected posterior priors are defined as

$$\pi_i^*(\theta_i) = \int \pi_i^N(\theta_i|z) m^*(z) \mathrm{d}z, \quad i = 1, \ldots, q,$$

where $m^*(z)$ is an arbitrary predictive density, which can be improper, for the imaginary training sample z. When M_1, say, is nested in the other models, Pérez & Berger (2002) propose $m^*(z) = m_1^N(z)$, and therefore this method applied to two nested-encompassing models produces the intrinsic priors introduced by Moreno *et al.* (1998).

1.1 Integral Priors

Integral priors have been introduced in Cano *et al.* (2008) for model selection when two models, M_i: $\{f_i(x|\theta_i),\ \theta_i \in \Theta_i\}$, $i = 1, 2$, are under consideration. These prior distributions, $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$, are defined as the solution of a system of integral equations, and can be considered as a generalisation of the expected posterior priors introduced in Pérez & Berger (2002). Concretely, integral priors $\{\pi_1(\theta_1),\ \pi_2(\theta_2)\}$ satisfy the integral equations

$$\pi_1(\theta_1) = \int \pi_1^N(\theta_1|z_1) m_2(z_1) \mathrm{d}z_1,$$

$$\pi_2(\theta_2) = \int \pi_2^N(\theta_2|z_2) m_1(z_2) \mathrm{d}z_2,$$

where $\pi_i^N(\theta_i)$ is a prior distribution for estimation (typically a noninformative prior), $\pi_i^N(\theta_i|z) \propto f_i(z|\theta_i)\pi_i^N(\theta_i)$, $m_i(z) = \int f_i(z|\theta_i)\pi_i(\theta_i)\mathrm{d}\theta_i$, $i = 1, 2$, and z_1 and z_2 are (minimal) imaginary training samples for θ_1 and θ_2, respectively. When θ_i or z_i are discrete, the corresponding integrals are replaced by sums with no loss of generality. These integral equations balance each model with respect to the other one since the prior $\pi_i(\theta_i)$ is derived from the marginal $m_j(z_i)$, and therefore from $\pi_j(\theta_j)$, $j \neq i$, as an unknown generalised expected posterior prior.

Integral priors can be used to compare both nested and non-nested models without the need to make this distinction, and the application of this methodology does not need to specify a predictive distribution as with the expected posterior priors approach.

Some examples where the exact form of integral priors is known are studied in Cano *et al.* (2008). The simplest one is the point null hypothesis testing $H_0: \theta = \theta^*$ *versus* $H_1: \theta \neq \theta^*$, for which the integral priors are $\pi_1(\theta_1) = \delta_{\theta^*}(\theta_1)$ and $\pi_2(\theta_2) = \int \pi_2^N(\theta_2|z) f_1(z|\theta^*)\mathrm{d}z$, respectively. A further example is the comparison of two location models, where the priors $\pi_i(\theta_i) = \pi_i^N(\theta_i) = 1$, $i = 1, 2$, are integral priors. This extends to the comparison of two scale and two location-scale models, the initial default priors, $\pi_i(\sigma_i) = \pi_i^N(\sigma_i) = 1/\sigma_i$, $i = 1, 2$, for the former, and $\pi_i(\mu_i, \sigma_i) = \pi_i^N(\mu_i, \sigma_i) = 1/\sigma_i$, $i = 1, 2$, for the latter, being integral priors. Further examples have been provided in Cano *et al.* (2008), while significant differences between intrinsic and integral priors have been discussed in Cano *et al.* (2018).

In general, closed-form solutions to the integral equations are unavailable. To deal with this drawback we established an equivalence between the solution of the integral equations and the convergence of two Markov chains. Specifically, Cano *et al.* (2008) considered Markov chains (θ_1^t) and (θ_2^t) whose σ-finite invariant *distributions* are the integral priors.

The transition $\theta_1' \to \theta_1$ of the Markov chain (θ_1^t) is the marginal of the following conditional steps:

1. Simulate a training sample $z_2 \sim f_1(z_2|\theta_1')$
2. Simulate $\theta_2 \sim \pi_2^N(\theta_2|z_2)$
3. Simulate a training sample $z_1 \sim f_2(z_1|\theta_2)$
4. Simulate $\theta_1 \sim \pi_1^N(\theta_1|z_1)$,

and the transition $\theta_2' \to \theta_2$ of the Markov chain (θ_2^t) is obtained in a symmetric manner, with both chains being simultaneously recurrent or transient. Therefore, the study of integral priors can be performed considering the stochastic stability of these Markov chains, as in Cano *et al.* (2007a),

Cano *et al.* (2007b), Cano & Salmerón (2013), Salmerón *et al.* (2015), Cano & Salmerón (2016), and Cano *et al.* (2018).

In some situations, the Markov chains attached to integral priors are positive recurrent, and even sometimes positive Harris recurrent, in which case the integral priors are proper prior distributions. This implies that the Bayes factor $B_{21}(x)$ is defined without ambiguity (Robert, 2001) and can be estimated by the Monte Carlo estimator

$$\hat{B}_{21}(x) = \frac{\sum_{t=1}^{T} f_2(x|\theta_2^t)}{\sum_{t=1}^{T} f_1(x|\theta_1^t)} \tag{1}$$

associated with the above Markov chains.

In the general case, Cano & Salmerón (2013) have proposed to consider constrained imaginary training samples z when the simulation of the chains is performed, which guarantees the existence and the uniqueness of invariant probability measures. This approach has been implemented successfully in several situations, see Cano & Salmerón (2013), and Cano *et al.* (2018). However, the simulation of constrained imaginary training samples can involve high computing time, specially when the accept-reject method is implemented. In this article, we propose a new approach that circumvents the need to simulate constrained imaginary training samples and that enables the methodology to be extended to comparing more than two models at once.

The plan of this paper is as follows. In Section 2, we generalise the definition of integral priors for more than two models. In Section 3, we propose to add copies of the models under consideration by introducing artificial compact parametric spaces, and we show that integral priors are proper priors. Section 4 is dedicated to show how the methodology applies to the problem of testing if the mean of a normal population is negative, zero, or positive. This multiple non-nested hypothesis testing problem has previously been studied in Berger & Mortera (1999) assuming known variance when using both intrinsic and fractional Bayes factor. However, as pointed out by Berger & Mortera (1999), these approaches do not correspond to genuine Bayes factors attached to specific prior distributions, and there is no a general methodology that, when applied to this model selection problem, produces priors with a satisfactory answer. In Section 5, we apply our methodology to the one way ANOVA model. Section 6 is dedicated to the problem of testing the Poisson *versus* the negative binomial using integral priors; a model selection problem where the definition of intrinsic priors presents a certain degree of arbitrariness when the initial default prior for estimation is the natural choice, that is, the Jeffreys priors. In Section 7, we show how the simulation from the integral posterior can be performed. Finally, in Section 8 we present some relevant conclusions and outline incoming research.

2 The Markov Chains Associated With Integral Priors

In this section, we extend the definition of the integral priors when comparing more than two models, through a generalisation of the above Markov chains.

As a starter, it proves convenient to represent the simulation of the dual chains (θ_1^t) and (θ_2^t) as shown in Figure 1, where each single step $\theta_i \rightarrow \theta_j$, $i \neq j$, is carried out by first simulating $z_j \sim f_i(z_j|\theta_i)$, which we call an imaginary training sample of sufficient size to make $\pi_j^N(\theta_j|z_j)$ proper, and second simulating $\theta_j \sim \pi_j^N(\theta_j|z_j)$, that is, simulating from the marginal transition

$$\pi_{ij}(\theta_i, \theta_j) = \int f_i(z_j|\theta_i)\pi_j^N(\theta_j|z_j)\mathrm{d}z_j. \tag{2}$$

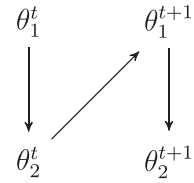

FIGURE 1. *Transition of the Markov chains for two models under comparison.*

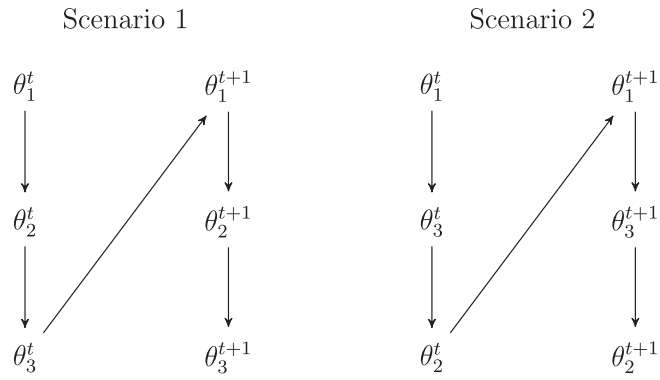

FIGURE 2. *Two competing scenarios to combine conditional updating steps when comparing three models.*

More specifically, given an initial point $\theta_1^1 \in \Theta_1$, we carry out the transitions $\theta_1^1 \to \theta_2^1 \to \theta_1^2 \to \theta_2^2 \ldots$ obtaining two marginal (interleaved) Markov chains, (θ_1^t) and (θ_2^t).

When starting with an initial point $\theta_2^1 \in \Theta_2$, the process maintains the same *stochastic behaviour*.

However, with more than two models under comparison, there exist several possibilities to combine simulations from $\pi_{ij}(\theta_i, \theta_j)$ in a sequence of simulation steps, towards obtaining Markov chains on each parametric space. Consider the case with three models. Figure 2 shows two scenarios to combine single steps. It is worth noting the similarity with Gibbs sampling. However, contrary to what happens with Gibbs sampling where the chosen order in which we carry out the simulation does not impact the stationary distribution (albeit it does impact the convergence rate), in the context we are dealing with, order matters and each scenario can produce a Markov chain on Θ_i with a different stationary distribution, $i = 1, 2, 3$. Suppose that M_1 corresponds to the point null hypothesis $\theta_1 = \theta_1^*$. Therein the stationary distribution $\pi_2(\theta_2)$ is $\pi_{12}(\theta_1^*, \theta_2)$ under scenario 1, and $\int \pi_{32}(\theta_3, \theta_2)\pi_{13}(\theta_1^*, \theta_3)\mathrm{d}\theta_3$ under scenario 2, which obviously differ. Hence the specific way in which the single steps are combined can produce different answers, with the consequent problem of defining integral priors for more than two models.

A straightforward proposal that makes this issue vanish is to randomly choose the ordering in which the steps are performed: if we have carried out the steps $\theta_{i_1}^t \to \theta_{i_2}^t \to \theta_{i_3}^t$ for a given sequence (i_1, i_2, i_3), then we select at random a sequence (j_1, j_2, j_3) with the condition that $j_1 \neq i_3$, and then we perform the steps $\theta_{i_3}^t \to \theta_{j_1}^{t+1} \to \theta_{j_2}^{t+1} \to \theta_{j_3}^{t+1}$, as shown in Figure 3.

This produces three Markov chains, (θ_1^t), (θ_2^t) and (θ_3^t), and leads to the following definition.

Definition *The multiple-model-comparison integral priors $\pi_1(\theta_1)$, $\pi_2(\theta_2)$ and $\pi_3(\theta_3)$, are the σ-finite invariant measures for the Markov chains (θ_1^t), (θ_2^t) and (θ_3^t), respectively, when using a random ordering of the conditional updates.*

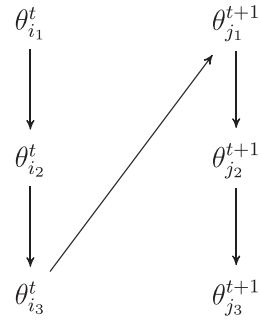

FIGURE 3. *Random ordering of the conditional updating steps when comparing three models:* (i_1, i_2, i_3) *and* (j_1, j_2, j_3) *are random permutations of* $(1, 2, 3)$, $j_1 \neq i_3$.

This definition reproduces the original one (with two models) and naturally extends to any number of models.

3 Proper Integral Priors

3.1 Auxiliary Compactified Models

As indicated above, for some model comparisons involving two models, the Markov chains associated with integral priors are recurrent Markov chains. In the general case, Cano & Salmerón (2013) have proposed to consider constrained imaginary training samples z when the simulation of the chains is performed, which guarantees both the existence and the uniqueness of invariant probability measures. In brief, their practical recommendation consists in keeping the imaginary training samples constrained within an interval of $\pm 5s$ about the sample mean, where s is the sample standard deviation. This approach has been implemented successfully in several situations comparing two models, see Cano & Salmerón (2013) and Cano *et al.* (2018). However, since the simulation of constrained imaginary training samples is carried out using some accept-reject method (for instance, a truncated normal distribution Robert, 1995), this induces a consequent computational overhead. The computational alternative developed in this article consists in including in the comparison an artificial copy of each model of interest but with a corresponding artificial compact parametric space.

Consider for instance the case of two models under comparison, $M_i{:}f_i(x|\theta_i),\ \theta_i \in \Theta_i,\ i = 1, 2$, and their attached improper priors $\pi_1^N(\theta_1)$ and $\pi_2^N(\theta_2)$, respectively. We then introduce an auxiliary model M_3 that is a copy of model M_1 with a parameter space restricted to a compact, that is,

$$M_3{:}f_1(x|\theta_3),\ \ \theta_3 \in \Theta_3,$$

where Θ_3 is an artificial compact subset of Θ_1, hence $\pi_3^N(\theta_3) \propto \pi_1^N(\theta_3)\mathbb{I}_{\Theta_3}(\theta_3)$ is a proper prior. Our goal is to obtain integral priors for both θ_1 and θ_2, but we can resort to our generalised methodology of constructing integral priors for the three models $\{M_1, M_2, M_3\}$, which produces Markov chains (θ_i^t) with transition densities $Q_i(\theta_i|\theta_i')$, and resulting integral priors $\pi_i(\theta_i)$, that is,

$$\pi_i(\theta_i) = \int Q_i(\theta_i|\theta_i')\pi_i(\theta_i')\mathrm{d}\theta_i',\ \ i = 1, 2, 3.$$

Note that we are using the conditional notation for the marginal transition density $Q_i(\theta_i|\theta_i')$, while the alternative notation $\pi_{ij}(\theta_i, \theta_j)$ corresponds for the between-model transition $\theta_i \rightarrow \theta_j,\ i \neq j$.

Since $\pi_3^N(\theta_3)$ is a proper prior, we need no imaginary training sample for θ_3. In this case,

$$\pi_{13}(\theta_1, \theta_3) = \pi_{23}(\theta_2, \theta_3) = \pi_3^N(\theta_3).$$

Thus, the Markov chain (θ_3^t) is a sequence of i.i.d. random variables, $\theta_3^t \sim \pi_3^N(\theta_3)$, and therefore $\pi_3^N(\theta_3)$ is the integral prior for θ_3.

The study of the stochastic stability of the corresponding Markov chains (θ_1^t) and (θ_2^t) entails a slightly greater level of complexity. The most natural approach is based on the 2-step transition densities

$$Q_i^2(\theta_i|\theta_i') = \int Q_i(\theta_i|\tilde{\theta}_i) Q_i(\tilde{\theta}_i|\theta_i') \mathrm{d}\tilde{\theta}_i, \quad i = 1, 2.$$

For $i = 1, 2$, both the transition density $Q_i(\theta_i|\theta_i')$ and the 2-step transition density $Q_i^2(\theta_i|\theta_i')$, are mixtures of different conditional densities due to the randomness in the order of the between-model sequences, as illustrated by Figure 4. Each component of the mixture $Q_i^2(\theta_i|\theta_i')$ corresponds to a specific realisation of the sequences (i_1, i_2, i_3), (j_1, j_2, j_3), and (k_1, k_2, k_3). However, all components of this mixture share the crucial property that a simulation from $\pi_3^N(\theta_3)$ occurs within the sequence, since $j = 3$ is necessarily part of the sequence (j_1, j_2, j_3). For instance, with regard to the mixture $Q_1^2(\theta_1|\theta_1')$, configuration 1 in Figure 5 corresponds to the component

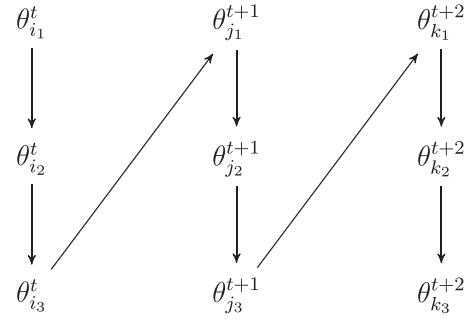

FIGURE 4. *Illustration of a 2-step transition when comparing three models:* (i_1, i_2, i_3), (j_1, j_2, j_3), *and* (k_1, k_2, k_3) *are random permutations of* $(1, 2, 3)$, $i_3 \neq j_1$, $j_3 \neq k_1$.

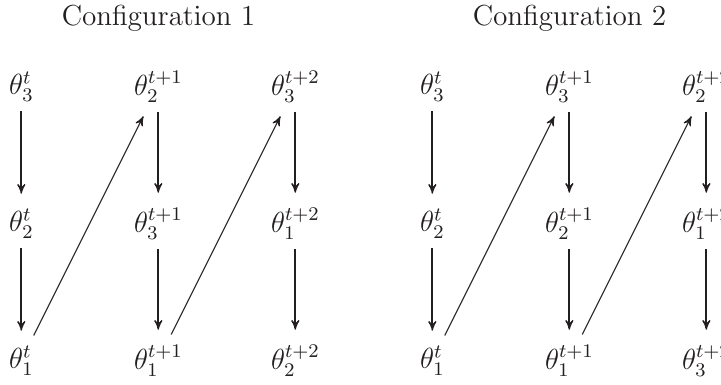

FIGURE 5. *Two realisations of the 2-step transition process given in Figure 4.*

$$\int \pi_3^N(\theta_3)\pi_{31}(\theta_3, \theta_1)\mathrm{d}\theta_3,$$

and configuration 2 corresponds to the component

$$\int \pi_3^N(\theta_3)\pi_{32}(\theta_3, \theta_2)\pi_{21}(\theta_2, \tilde{\theta}_1)\pi_{12}(\tilde{\theta}_1, \theta_2')\pi_{21}(\theta_2', \theta_1)\mathrm{d}\theta_2'\mathrm{d}\tilde{\theta}_1\mathrm{d}\theta_2\mathrm{d}\theta_3.$$

More precisely, it is straightforward to derive that all components of these mixtures are necessarily equal to one of four cases that can be schematically represented as follows:

$$1\to3\to1,\ \ 1\to3\to2\to1,$$

$$1\to3\to1\to2\to1,\ \text{ and } 1\to3\to2\to1\to2\to1.$$

Then $Q_1^2(\theta_1|\theta_1')$ does not depend on θ_1', and hence, $\pi_1(\theta_1) = Q_1^2(\theta_1|\theta_1')$ is the integral prior for θ_1:

$$\begin{aligned}\int \pi_i(\theta_i)Q_i(\theta_i^*|\theta_i)\mathrm{d}\theta_i &= \int Q_i^2(\theta_i|\theta_i')Q_i(\theta_i^*|\theta_i)\mathrm{d}\theta_i \\ &= \int Q_i(\theta_i|\tilde{\theta}_i)Q_i(\tilde{\theta}_i|\theta_i')Q_i(\theta_i^*|\theta_i)\mathrm{d}\theta_i\mathrm{d}\tilde{\theta}_i \\ &= \int Q_i(\tilde{\theta}_i|\theta_i')Q_i^2(\theta_i^*|\tilde{\theta}_i)\mathrm{d}\tilde{\theta}_i = Q_i^2(\theta_i^*|\theta_i') = \pi_i(\theta_i^*).\end{aligned}$$

Similarly, the 2-step transition for (θ_2') is the integral prior for θ_2, that is, $\pi_2(\theta_2) = Q_2^2(\theta_2|\theta_2')$. From a practical perspective, we point out that all simulation steps that precede the latest simulation of θ_3 in $Q_i(\theta_i|\theta_i')$ are superfluous since θ_3 is generated independently.

However, these resulting integral priors do depend on which model is copied into a compactified version. Towards the elimination of the resulting arbitrariness, we propose to duplicate *each* model into a compactified version. For instance, when facing two models, we add to model M_2 the model

$$M_4{:}f_2(x|\theta_4),\ \ \theta_4 \in \Theta_4,$$

where Θ_4 is an artificial compact subset of Θ_2 with $\pi_4^N(\theta_4) \propto \pi_2^N(\theta_4)\mathbb{I}_{\Theta_4}(\theta_4)$ thus a proper prior. While one could consider comparing $\{M_3, M_4\}$ instead of $\{M_1, M_2\}$, since the corresponding Bayes factor $B_{34}(x)$ is then well-defined, the models of interest remain M_1 and M_2. In fact, in the case where M_1 corresponds to the point null hypothesis $\theta_1 = \theta_1^*$, the Bayes factor for the comparison M_3 *versus* M_4 is

$$B_{34}(x) = \frac{f_1(x|\theta_1^*)\int_{\Theta_4}\pi_2^N(\theta_2)\mathrm{d}\theta_2}{\int_{\Theta_4} f_2(x|\theta_2)\pi_2^N(\theta_2)\mathrm{d}\theta_2},$$

which typically goes to $+\infty$ as Θ_4 approaches Θ_2. Therefore, moving to the set of models $\{M_1, M_2, M_3, M_4\}$ is purely procedural and solely intended to derive a well-defined integral prior on both M_1 and M_2.

For the same reason as above, the training sample size for θ_4 is zero and the integral prior is again $\pi_4^N(\theta_4)$. The 2-step transition $Q_i^2(\theta_i|\theta_i')$ does not depend on θ_i', and hence, $\pi_i(\theta_i) = Q_i^2(\theta_i|\theta_i')$ is the integral prior for $\theta_i,\ i = 1, 2$. Besides, $Q_1^2(\theta_1|\theta_1')$ is a mixture of densities and

the components of this mixture are reduced to eight cases that can be schematically represented as follows:

$$1\to3\to1,\ \ 1\to3\to2\to1,$$

$$1\to3\to1\to2\to1,\ \ 1\to3\to2\to1\to2\to1,$$

$$1\to4\to1,\ \ 1\to4\to2\to1,$$

$$1\to4\to1\to2\to1,\ \text{and}\ 1\to4\to2\to1\to2\to1,$$

which means that there exist functions $H_1^1(\theta_1, \theta_3)$ and $H_1^2(\theta_1, \theta_4)$, defined for $\theta_1, \theta_3 \in \Theta_1$ and $\theta_4 \in \Theta_2$, such that the integral prior $\pi_1(\theta_1)$ is

$$\pi_1(\theta_1) = Q_1^2(\theta_1|\theta_1') = \int \pi_3^N(\theta_3) H_1^1(\theta_1, \theta_3)\mathrm{d}\theta_3 + \int \pi_4^N(\theta_4) H_1^2(\theta_1, \theta_4)\mathrm{d}\theta_4.$$

Similarly, there exist functions $H_2^1(\theta_2, \theta_3)$ and $H_2^2(\theta_2, \theta_4)$, defined for $\theta_2, \theta_4 \in \Theta_2$ and $\theta_3 \in \Theta_1$, such that

$$\pi_2(\theta_2) = Q_2^2(\theta_2|\theta_2') = \int \pi_3^N(\theta_3) H_2^1(\theta_2, \theta_3)\mathrm{d}\theta_3 + \int \pi_4^N(\theta_4) H_2^2(\theta_2, \theta_4)\mathrm{d}\theta_4,$$

is the integral prior for θ_2.

Obviously, this device of adding copycat models with an artificial compact parametric space applies to more than two models. We start from $q \geq 2$ models

$$M_i{:}f_i(x|\theta_i),\ \ \theta_i \in \Theta_i,\ \ i = 1, \ldots, q,$$

under comparison and associated improper prior distributions $\pi_i^N(\theta_i)\, i = 1, \ldots, q$. Given a compact subset $\Theta_{q+i} \subseteq \Theta_i$ such that $\int_{\Theta_{q+i}} \pi_i^N(\theta_i)\mathrm{d}\theta_i < +\infty$, we consider a copy of M_i but with parameter space Θ_{q+i}, that is,

$$M_{q+i}{:}f_i(x|\theta_{q+i}),\ \ \theta_{q+i} \in \Theta_{q+i},$$

and $\pi_{q+i}^N(\theta_{q+i}) \propto \pi_i^N(\theta_{q+i})\mathbb{I}_{\Theta_{q+i}}(\theta_{q+i}),\ i = 1, \ldots, q$. Then integral priors for $\{M_1, \ldots, M_q, \ldots, M_{2q}\}$ do exist, they all are unique and proper priors, and the integral prior for θ_i satisfies

$$\pi_i(\theta_i) = Q_i^2(\theta_i|\theta_i') = \sum_{j=1}^{q} \int \pi_{q+j}^N(\theta_{q+j}) H_i^j(\theta_i, \theta_{q+j})\mathrm{d}\theta_{q+j},$$

where the function $H_i^j(\theta_i, \theta_{q+j})$ is defined for $\theta_i \in \Theta_i$ and $\theta_{q+j} \in \Theta_j,\ i, j = 1, \ldots, q$.

3.2 Selecting the Compact Sets

Without loss of generality consider two models under comparison, M_1 and M_2, with their copies over compact parametric spaces, M_3 and M_4, respectively. It is quite sensible to seek compact sets such that the posterior probabilities

$$P_x^3 = \int_{\Theta_3} \pi_1^N(\theta_1|x)\mathrm{d}\theta_1 \text{ and } P_x^4 = \int_{\Theta_4} \pi_2^N(\theta_2|x)\mathrm{d}\theta_2$$

are close to one. While there are a variety of ways to choose these compact sets, we opt for the simple solution where both Θ_3 and Θ_4 are the Cartesian products of compact intervals based on marginal posterior quantiles in such a way that the compact sets include the highest posterior values. Note that here the posterior distributions are associated with the original prior distributions, independently for each model. More specifically, our proposal is made of credible intervals with posterior probability $1 - \alpha$, based on the quantiles $\alpha/2$ and $1 - \alpha/2$, with $\alpha \in (0, 1)$ small, typically $\alpha \leq 0.05$. Note that the computation of such HPD regions usually requires MCMC samples. This differs from the costlier requirement of simulating constrained imaginary training samples in the Markov chains that define integral priors under the approach proposed in Cano & Salmerón (2013). The following example illustrates the behaviour of this method.

Example *For the data $x = (x_1, \ldots, x_n)$, consider the models*

$$M_1 : f_1(x_j|\theta_1) = \frac{1}{\theta_1}\exp(-x_j/\theta_1),\ j = 1, \ldots, n,\ \theta_1 \in (0, 1),$$

and

$$M_2 : f_2(x_j|\theta_2) = \frac{1}{\theta_2}\exp(-x_j/\theta_2),\ j = 1, \ldots, n,\ \theta_2 \in (1, +\infty),$$

with $\pi_i^N(\theta_i) = 1/\theta_i,\ i = 1, 2$. Note that $\sum_j x_j$ is a sufficient statistic for θ, and that given prior distributions for $\theta_i,\ i = 1, 2$, the Bayes factor is determined by n and $\bar{x}$. For a sample (see supplementary material) of size $n = 15$ and mean 0.7, we implement the above methodology of integral priors, thus defining the collection of models $\{M_1, M_2, M_3, M_4\}$, where $\Theta_{i+2} = [a_{i+2}, b_{i+2}]$ is a $100(1 - \alpha)\%$ credible interval based on the quantiles $\alpha/2$ and $1 - \alpha/2$ of $\pi_i^N(\theta_i|x),\ i = 1, 2$. We also opt for using training samples z of size one for both θ_1 and θ_2. Simulation from the posterior distributions is straightforward:

$$\theta_1 = 1/(1 - z^{-1}\log u_1),\ u_1 \sim \mathcal{U}(0, 1),$$

and

$$\theta_2 = -z/\log(u_2(1 - e^{-z}) + e^{-z}),\ u_2 \sim \mathcal{U}(0, 1).$$

For $\alpha \in \{0.05, 0.01, 0.005, 0.001\}$, Table 1 provides $\hat{B}_{12}(x) = 1/\hat{B}_{21}(x)$ where $\hat{B}_{21}(x)$ is given by (1), the approximate values of $B_{12}(x)$, as well as the estimated posterior probability of model

Table 1. Estimation of the Bayes factor $B_{12}(x)$ and posterior probability $P(M_1|x)$, based on 10^6 simulations from the integral priors.

α	a_3	b_3	a_4	b_4	$\hat{B}_{12}(x)$	$\hat{P}(M_1|x)$
0.05	0.44	0.97	1.00	1.62	8.98	0.90
0.01	0.39	0.99	1.00	1.90	9.10	0.90
0.005	0.37	1.00	1.00	2.02	9.09	0.90
0.001	0.34	1.00	1.00	2.28	9.16	0.90

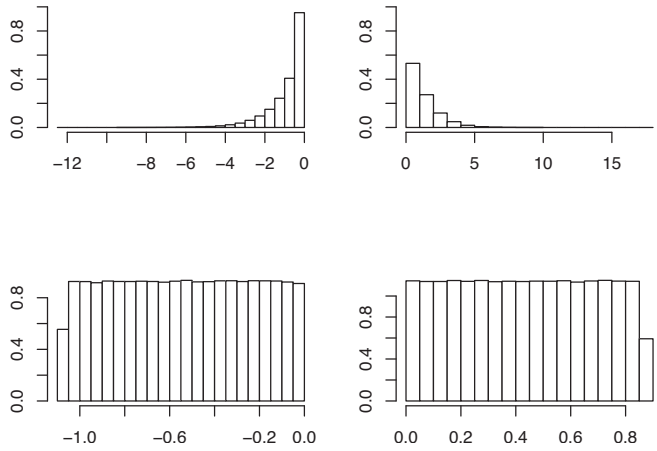

FIGURE 6. *Histograms of 10^6 simulations from the integral priors for* $\log\theta_i$, $i = 1, 2$ *(first row:* $\log\theta_1$ *left,* $\log\theta_2$ *right), and* $i = 3, 4$ *(second row:* $\log\theta_3$ *left,* $\log\theta_4$ *right).*

M_1 if the prior probabilities of M_1, M_2, M_3, and M_4 are 1/2, 1/2, 0, and 0, respectively. Hence, $\hat{P}(M_1|x) = \hat{B}_{12}(x)/(1 + \hat{B}_{12}(x))$.

It is worth noting that under the unconstrained Exponential model

$$f(x_j|\theta) = \theta^{-1}\exp(-x_j/\theta),\ \ j = 1\ldots, n,\ \ \theta \in (0, +\infty),$$

and $\pi^N(\theta) \propto 1/\theta$, the posterior probability of $\Theta_1 = (0, 1)$ is 0.89.

The histograms corresponding to the chains $(\log\theta_i^t)$, $i = 1, 2, 3, 4$, using the associated Markov chains with the compact sets for $\alpha = 0.001$ (last row in Table 1), are shown in Figure 6.

To evaluate the Monte Carlo variability in Table 1, we repeated 100 times the estimations, that is, we simulated the Markov chains 100 times and computed the corresponding estimator $\hat{B}_{12}(x)$. The standard deviation of the 100 values of $\hat{B}_{12}(x)$ was 0.02 for each value of α.

The intrinsic priors for the arithmetic intrinsic Bayes factor do not exist for this example because the expectation of $B_{12}^N(z)$ is not finite. The encompassing approach for intrinsic priors has been suggested in such situations. This approach proposes embedding M_1 and M_2 in the larger model M_0: $\theta \in (0, +\infty)$, and then to obtain intrinsic priors for nested models to compute $B_{20}^I(x)$ and $B_{01}^I(x)$, and finally $B_{21}^I(x)$ is defined as $B_{20}^I(x)B_{01}^I(x)$. Moreno (2005) has stated that this approach does not work for this example.

Instead, Moreno (2005) has proposed an ad hoc solution inspired from intrinsic priors for two different models: the point null hypothesis $\theta = 1$ and the unconstrained model $\theta > 0$. Then the proposal is to consider the restriction of the intrinsic prior for the unconstrained model to $\theta < 1$ and $\theta > 1$ to obtain two priors for the computation of the Bayes factor $\tilde{B}_{12}^I(x)$. For a sample x of size $n = 15$ and mean 0.7, this Bayes factor is $\tilde{B}_{12}^I(x) = 7.6$, which is similar to the values in Table 1. Berger & Mortera (1999) have obtained other priors for the comparison M_1 *versus* M_2; for example, the intrinsic priors for the median intrinsic Bayes factor and for the fractional Bayes factor. With these priors the resulting Bayes factors are $\tilde{B}_{12}^{MIBF}(x) = 5.6$ and $\tilde{B}_{12}^F(x) = 3.1$, respectively. Both values are significantly different from those obtained with our approach. This is partly explained by the fact that $\int_1^{+\infty}\pi_2^I(\theta_2)\mathrm{d}\theta_2/\int_0^1\pi_1^I(\theta_1)\mathrm{d}\theta_1$ is 2.67 for the priors used in

$\tilde{B}_{12}^{F}(x)$, and 1.46 for the priors used in $\tilde{B}_{12}^{MIBF}(x)$, and this translates into an a priori advantage for model M_2, as pointed out in Berger & Mortera (1999).

4 Test on a Normal Mean Under Dependent Samples

In the specific context of paired samples in which each subject is measured twice, resulting in pairs of observations, there are three natural hypotheses regarding the mean difference, μ, namely, $\mu < 0$, $\mu = 0$, and $\mu > 0$. This multiple non-nested hypothesis testing problem has been previously studied in Berger & Mortera (1999), assuming known variance and comparing the encompassing arithmetic intrinsic Bayes factor (EIBF), the encompassing expected intrinsic Bayes factor (EEIBF), the median intrinsic Bayes factor, and the fractional Bayes factor. Unfortunately, these approaches do not correspond to Bayes factors with respect to genuine prior distributions (see Berger & Mortera (1999), p. 552). In this section, we therefore apply the theory of integral priors for this testing problem.

Under model M_i the difference data $x = (x_1, \ldots, x_n)$ are independently drawn from the Normal distribution $\mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, 3$, with $\mu_1 < 0$, $\mu_2 = 0$, and $\mu_3 > 0$, respectively, that is,

$$f_i(x|\mu_i, \sigma_i) = (2\pi)^{-n/2}\sigma_i^{-n}\exp\left(-\frac{1}{2\sigma_i^2}\sum_{j=1}^{n}(x_j - \mu_i)^2\right), \quad i = 1, 3,$$

$$f_2(x|\sigma_2) = (2\pi)^{-n/2}\sigma_2^{-n}\exp\left(-\frac{1}{2\sigma_2^2}\sum_{j=1}^{n}x_j^2\right),$$

and we consider the default estimation priors $\pi_i^N(\theta_i) = 1/\sigma_i$, $i = 1, 2, 3$, where $\theta_1 = (\mu_1, \sigma_1)$, $\theta_2 = \sigma_2$, and $\theta_3 = (\mu_3, \sigma_3)$.

To investigate this setting, we produced a simulated sample x of size $n = 15$ from the Normal distribution with mean -1 and variance 4. The observed sample mean and standard deviation were -0.927 and 2.530, respectively. We thus consider copies M_4, M_5, and M_6, with compact parametric spaces Θ_i, $i = 4, 5, 6$, of the form

$$\theta_4 = (\mu_4, \sigma_4) \in \Theta_4 = [a_4, b_4] \times [c_4, d_4],$$

$$\theta_5 = \sigma_5 \in \Theta_5 = [c_5, d_5],$$

and

$$\theta_6 = (\mu_6, \sigma_6) \in \Theta_6 = [a_6, b_6] \times [c_6, d_6],$$

respectively.

For models M_1 and M_3, imaginary training samples are samples, z, of size 2, and the corresponding posterior distributions are

$$\pi_i^N(\mu_i, \sigma_i|z) \propto \sigma_i^{-3}\exp\left(-\frac{s_z^2}{2\sigma_i^2}\right)\exp\left(-\frac{(\bar{z} - \mu_i)^2}{\sigma_i^2}\right), \quad i = 1, 3,$$

where $\bar{z}$ and $s_z \neq 0$ are the sample mean and standard deviation of z, respectively. The posterior distribution $\pi_i^N(\mu_i|z)$ is proportional to

$$\int_0^{+\infty} \sigma_i^{-3} \exp\left(-\frac{s_z^2}{2\sigma_i^2}\right) \exp\left(-\frac{(\bar{z}-\mu_i)^2}{\sigma_i^2}\right) d\sigma_i = \frac{1}{s_z^2 + 2(\bar{z}-\mu_i)^2},$$

an half-Cauchy density, and therefore the simulation from $\pi_i^N(\mu_i|z)$ can be carried out using the probability integral transform method, that is, we can generate $u,\ v \sim \mathcal{U}(0,1)$ and take

$$\mu_1 = \bar{z} - \frac{s_z}{\sqrt{2}} \cot\left(\frac{u}{2}\left(\pi - 2\arctan\left(\sqrt{2}\bar{z}/s_z\right)\right)\right)$$

and

$$\mu_3 = \bar{z} - \frac{s_z}{\sqrt{2}} \cot\left(\frac{v-1}{2}\left(\pi + 2\arctan\left(\sqrt{2}\bar{z}/s_z\right)\right)\right).$$

To simulate from $\pi_i^N(\sigma_i|\mu_i, z)$ we can generate ξ_i from the Exponential distribution $\mathcal{E}xp(1)$ and take

$$\sigma_i = \sqrt{s_z^2 + 2\frac{(\bar{z}-\mu_i)^2}{2\xi_i}},\ \ i = 1, 3.$$

For model M_2, imaginary training samples are samples of size $1, z \neq 0$, and the posterior distribution is

$$\pi_2^N(\sigma_2|z) \propto \sigma_2^{-2} \exp\left(-\frac{z^2}{2\sigma_2^2}\right).$$

Therefore, to simulate from $\pi_2^N(\sigma_2|z)$ we can generate $\xi_2 \sim \mathcal{G}a(1/2, 1)$ and take $\sigma_2 = |z|/\sqrt{2\xi_2}$.

For $\alpha \in \{0.05, 0.01,\ 0.005\}$ we again select the compact intervals as the $100(1-\alpha)\%$ credible intervals based on the quantiles $\alpha/2$ and $1-\alpha/2$ of the marginal posterior distribution of $\pi_i^N(\theta_i|x),\ i = 1, 2, 3$, that is,

$$\alpha/2 = \int_{-\infty}^{a_i+3} \pi_i^N(\mu_i|x) d\mu_i,\ \ 1-\alpha/2 = \int_{-\infty}^{b_i+3} \pi_i^N(\mu_i|x) d\mu_i\ i = 1, 3,$$

$$\alpha/2 = \int_0^{c_i+3} \pi_i^N(\sigma_i|x) d\sigma_i,\ \text{ and } 1-\alpha/2 = \int_0^{d_i+3} \pi_i^N(\sigma_i|x) d\sigma_i\ i = 1, 2, 3.$$

Based on such compact sets, we apply the methodology of integral priors with 500,000 simulations in order to approximate the posterior probabilities of models M_1, M_2 and M_3 when selecting as set of prior probabilities $\{1/3, 1/3, 1/3, 0,\ 0, 0\}$. The approximated posterior probabilities of models M_1, M_2, and M_3 are shown in Table 2, again with very little variability when α decreases. The execution time was 20 seconds for each row in Table 2. On the other hand, we

Table 2. *Estimation of the posterior probabilities of models M_1, M_2, and M_3 in the Normal mean difference example.*

α	$\hat{P}(M_1\|x)$	$\hat{P}(M_2\|x)$	$\hat{P}(M_3\|x)$
0.05	0.51	0.42	0.06
0.01	0.50	0.43	0.06
0.005	0.50	0.44	0.06

have applied integral priors for $\{M_1, M_2, M_3\}$ with constrained imaginary training samples as proposed in Cano & Salmerón (2013) with Markov chains of length 500,000. As could be expected, the execution time was higher with this approach. Specifically, the execution time was 120 seconds and the posterior probabilities of models M_1, M_2, and M_3 were 0.52, 0.46, and 0.02, respectively.

For this example we have computed the fractional Bayes factors and the median intrinsic Bayes factors. The corresponding posterior probabilities of models $\{M_1, M_2, M_3\}$ are $\{0.41, 0.53, 0.06\}$ and $\{0.28, 0.63, 0.09\}$, respectively. Note that the conclusion from these methodologies would be to choose M_2, while the conclusion based on integral priors would be to choose M_1. However, despite this slight numerical difference, the posterior probability is not sufficiently conclusive in favour of either M_1 or M_2 in any case. In fact, although the fractional Bayes factor (0.77) and the median intrinsic Bayes factor (0.44) provide evidence against M_1, and the Bayes factor for integral priors (0.88) provides evidence against M_2, all of these pieces of evidence are not worth more than a bare mention, according to Jeffreys' scale of evidence.

Figure 7 shows the histograms and the empirical autocorrelations based on the simulations from the integral prior for $\log(-\mu_1)$ and $\log(\mu_3)$, corresponding to $\alpha = 0.005$.

5 Integral Priors for ANOVA

5.1 Original and Auxiliary Models

In the classical ANOVA setting, we now consider a sample x_i from the Normal population $\mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \ldots, k$, in order to handle on the hypothesis test

$$H_0: \mu_1 = \ldots = \mu_k$$
$$H_1: \mu_i \text{ are not all equal.}$$

Hence we are again dealing with a comparison between two models. Under model M_1 the k populations have mean μ and variance σ_1^2, and under model M_2 the population i has mean μ_i and variance σ_2^2, $i = 1, \ldots, k$. The default (estimation) prior distributions in this setting are $\pi_1^N(\mu, \sigma_1) = 1/\sigma_1$ and $\pi_2^N(\mu_1, \ldots, \mu_k, \sigma_2) = 1/\sigma_2$. Unfortunately, these priors are improper

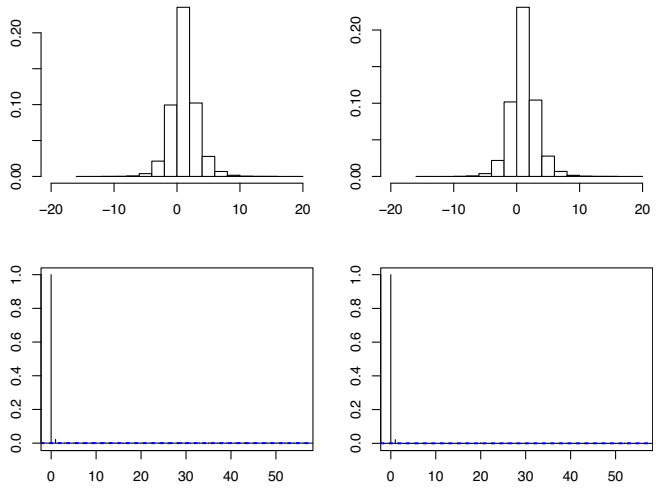

FIGURE 7. *Histograms and empirical autocorrelations based on simulations from the integral priors for* $\log(-\mu_1)$ *and* $\log(\mu_3)$ *in the Normal mean difference example.*

and therefore their use is not validated for model selection. In order to construct integral priors for ANOVA, we again consider copies M_3 and M_4 with compact parametric spaces.

The imaginary training samples are generated as follows:

For $\theta_1 \to \theta_2$. Given $\theta_1 = (\mu, \sigma_1)$

1. $j \sim \mathcal{U}\{1, \ldots, k\}$
2. $z_i \sim N(\mu, \sigma_1^2)$, $i \neq j$, and $z_j = (z_j^1, z_j^2)$, where $z_j^1, z_j^2 \sim N(\mu, \sigma_1^2)$
3. $z = (z_1, \ldots, z_k)$ is the imaginary training sample

For $\theta_2 \to \theta_1$. Given $\theta_2 = (\mu_1, \ldots, \mu_k, \sigma_2)$

1. $j_1, j_2 \sim \mathcal{U}\{1, \ldots, k\}$
2. $z_{j_1} \sim N(\mu_{j_1}, \sigma_2^2)$, $z_{j_2} \sim N(\mu_{j_2}, \sigma_2^2)$
3. $z = (z_{j_1}, z_{j_2})$ is the imaginary training sample

For model M_1, the sample size is 2, that is, $x = (z_1, z_2)$. The simulation from $\pi_1^N(\mu, \sigma_1|x)$ can be carried out as follows:

1. $\xi \sim \mathcal{G}a(1/2, 1)$, $\sigma_1 = \dfrac{|z_1 - z_2|}{2\sqrt{\xi}}$
2. $\mu \sim N(\overline{x}, \sigma_1^2/2)$.

For model M_2, the sample size is $n_j = 2$ for some $j \in \{i, \ldots, k\}$ and $n_i = 1$ for $i \neq j$. Then

$$\pi_2^N(\sigma_2|x) \propto \sigma_2^{-2} \exp\left(-\frac{s_j^2}{2\sigma_2^2}\right),$$

and therefore the simulation from $\pi_2^N(\mu_1, \ldots, \mu_k, \sigma_2|x)$ can be carried out as follows:

1. $\xi \sim \mathcal{G}a(1/2, 1)$, $\sigma_2 = \dfrac{|z_1 - z_2|}{2\sqrt{\xi}}$, where $x_j = (z_1, z_2)$
2. $\mu_i \sim N(\overline{x}^i, \sigma_2^2/n_i)$, $i = 1, \ldots, k$.

5.2 Example

To illustrate our methodology we consider $k = 3$ populations. The compact parametric spaces are chosen as Cartesian products

$$\Theta_3 = [a_3, b_3] \times [c_3, d_3]$$

and

$$\Theta_4 = [a_{41}, b_{41}] \times [a_{42}, b_{42}] \times [a_{43}, b_{43}] \times [c_4, d_4].$$

For $\alpha \in \{0.05, 0.01, 0.005\}$, we select the intervals as $100(1 - \alpha)\%$ credible intervals. For sample sizes $n_1 = n_2 = n_3 = 10$, we have considered sample means $\overline{x}^1 = \overline{x}^2 = 0$, $\overline{x}^3 \in \{0, 0.75, 1, 1.5\}$, and sample standard deviations $s_1 = s_2 = s_3 = 1$. As in previous examples, the compact parametric spaces are based on the posterior quantiles $\alpha/2$ and $1 - \alpha/2$ of the marginal posterior distribution obtained with $\pi_i^N(\theta_i)$, $i = 1, 2$. Table 3 produces estimated posterior probability $\hat{P}(M_1|x)$ under the prior probabilities $P(M_1) = P(M_2) = 1/2$ and $P(M_3) = P(M_4) = 0$.

Table 3. *Estimations of the posterior probabilities of model M_1 in the ANOVA example.*

	$\overline{x}^3$			
	0	0.75	1	1.5
$\alpha = 0.05$	0.93	0.70	0.38	0.03
$\alpha = 0.01$	0.94	0.70	0.41	0.03
$\alpha = 0.005$	0.94	0.71	0.41	0.03

The p-values for the ANOVA test were 1, 0.173, 0.051, and 0.003, for $\overline{x}^3 = 0, 0.75, 1$, and 1.5, respectively. Our reference priors are the g-priors as implemented in the R package BayesFactor with the three possible options for the scale of the inverse-chi-square priors for the g's: *medium*, *wide*, and *ultrawide*. The posterior probability of model M_1 were 0.82 (*medium*), 0.88 (*wide*), and 0.93 (*ultrawide*) when $\overline{x}^3 = 0$; 0.59 (*medium*), 0.65 (*wide*), and 0.73 (*ultrawide*) when $\overline{x}^3 = 0.75$; 0.38 (*medium*), 0.41 (*wide*), and 0.50 (*ultrawide*) when $\overline{x}^3 = 1$; and 0.06 (*medium*), 0.06 (*wide*), and 0.06 (*ultrawide*) when $\overline{x}^3 = 1.5$.

6 Poisson and Negative Binomial

The article by Moreno *et al.* (2021) is dedicated to the test of a Poisson model *versus* the negative binomial family, using intrinsic priors. In this section, we develop integral priors and compare the methodologies.

Let us consider the Poisson model

$$M_1 : f_1(x|\theta_1) = \frac{\theta_1^x e^{-\theta_1}}{x!}, \ \theta_1 > 0,$$

and the negative binomial models

$$M_r : f_r(x|\theta_r) = \binom{x + r - 2}{x} \theta_r^{r-1} (1 - \theta_r)^x, \ \theta_r \in (0, 1),$$

$r = 2, \ldots, q$. Jeffreys priors are $\pi_1^N(\theta_1) = \theta_1^{-1/2}$ and $\pi_r^N(\theta_r) = \theta_r^{-1}(1 - \theta_r)^{-1/2}$, $r = 2, \ldots, q$, and the minimal training sample size is one in all cases. With these priors, the posteriors are Gamma and Beta distributions, respectively:

$$\pi_1^N(\theta_1|z) = \mathcal{G}a(\theta_1|z + 1/2, 1)$$

and

$$\pi_r^N(\theta_r|z) = \mathcal{B}e(\theta_r|r - 1, z + 1/2), \ r \geq 2,$$

for $z \geq 0$ a minimal training sample.

As is pointed out in (Moreno *et al.*, 1998, pp. 1457–1458), the intrinsic priors methodology is not well defined when comparing Poisson *versus* geometric (M_2) because of the arbitrariness in the limiting procedure for defining the intrinsic priors and the corresponding Bayes factor. This issue can be *adjusted* if instead of using Jeffreys priors as the initial default estimation priors—which would be the natural choice—one uses Jaynes' priors, for which a minimal training sample is any value of x other than $x = 0$

Table 4. *The word 'may' per block in papers by James Madison.*

X (number of occurrences)	0	1	2	3	4	5	6
Frequency (number of blocks)	156	63	29	8	4	1	1

Here we illustrate the integral priors methodology with the data in Table 4 referring the number of times that the word 'may' appears per block in papers by James Madison, see Conigliani *et al.* (2000).

We have considered $q = 15$ and, as in previous examples, the compact parametric spaces are based on the quantiles $\alpha/2$ and $1 - \alpha/2$ of the posterior distributions $\pi_i^N(\theta_i|y),\ i = 1, \ldots, q$.

Based on such compact sets, we apply the methodology of integral priors with 50,000 simulations in order to approximate the posterior probabilities of models $M_1, \ldots, M_{15}$. The prior probabilities were 1/15 for models $M_1, \ldots, M_{15}$, and 0 for the artificial models. Using $\alpha = 0.05$, the approximated posterior probabilities where 0.68 for the geometric model M_2, 0.26 for M_3, 0.04 for M_4, and less than 0.01 for other models. The results were virtually identical with $\alpha = 0.001$.

For the purpose of comparing the methodologies, we have applied the approach proposed in Moreno *et al.* (2021) obtaining quite similar values for the posterior probabilities: 0.70 for M_2, 0.24 for M_3, 0.04 for M_4, 0.01 and lower for the other models.

Furthermore, we have used simulated data to compare the models $\{M_1, M_2, M_3, M_4, M_5\}$ using integral priors ($\alpha = 0.05$, and chains of length 50,000) and the intrinsic priors developed in Moreno *et al.* (2021). A sample of size n is simulated from the *True* model 100 times, and for each sample we have computed the posterior probabilities of the models, and the probability of the models based on Akaike weights. Table 5 displays the means of these 100 probabilities. It can be observed that the values are very similar.

7 Integral Posteriors and Evidence Approximation

In case copycat models with compact parametric spaces have been added, a simple Monte Carlo estimator of the Bayes factor for integral priors is based on the Markov chains (θ_i^t) associated with integral priors $\pi_i(\theta_i),\ i = 1, \ldots, q$. For example, we can estimate the Bayes factor $B_{21}(x)$ by $\sum_{t=1}^{T} f_2(x|\theta_2^t) / \sum_{t=1}^{T} f_1(x|\theta_1^t)$. However, it is well known that this is not the optimal strategy when the integral posterior $\pi_i(\theta_i|x) \propto \pi_i(\theta_i) f_i(x|\theta_i)$ is concentrated relative to the integral prior $\pi_i(\theta_i)$. In such situations general methods based on the integral posterior could produce estimates of the evidence, and therefore of the Bayes factor, in a more efficient way. Besides, the simulation from the integral posterior entails some additional difficulties because the integral prior is not known analytically. To overcome this issue we resort to a completion of the integral posterior as follows.

The simulation of θ_i^{t+1} given θ_i^t is carried out in several steps. The last two steps correspond to the simulation of a training sample z_i^{t+1}, and the simulation θ_i^{t+1} from $\pi_i^N(\theta_i|z_i^{t+1})$. Besides, since the 2-step transition density $Q_i^2(\theta_i|\theta_i')$ does not depend on θ_i', it follows that given any initial point θ_i^0, the subsequences $\{\theta_i^2,\ \theta_i^4,\ \theta_i^6,\ \ldots\}$ and $\{\theta_i^3,\ \theta_i^5,\ \theta_i^7,\ \ldots\}$ are iid from $\pi_i(\theta_i)$. Note that θ_i^t and θ_i^{t+1} are independent if any compact space is involved along the transition $\theta_i^t \to \theta_i^{t+1}$. In general, we consider the subsequence $(\check{\theta}_i^t)$ of the Markov chain (θ_i^t) such that $\check{\theta}_i^t$ and $\check{\theta}_i^{t+1}$ are independent, and the corresponding subsequence $(\check{z}_i^t)$ of (z_i^t). This sequence is the key ingredient of the completion method. If $p_i(z_i)$ is the stationary distribution of (z_i^t), then $(\check{z}_i^t)$ are iid from $p_i(z)$, and

Table 5. *Mean of the posterior probabilities using integral priors, intrinsic priors, and mean of the probabilities of the models using Akaike weights.*

	True model: M_1 with $\theta_1 = 0.5$														
	Integral priors					Intrinsic priors					Akaike weights				
n	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
15	0.25	0.15	0.19	0.20	0.21	0.23	0.17	0.19	0.20	0.21	0.24	0.15	0.19	0.21	0.21
50	0.29	0.10	0.18	0.21	0.23	0.29	0.10	0.17	0.21	0.23	0.31	0.09	0.17	0.21	0.23
100	0.35	0.06	0.15	0.21	0.24	0.37	0.04	0.14	0.20	0.24	0.38	0.04	0.14	0.20	0.24
300	0.59	0.00	0.05	0.14	0.22	0.59	0.00	0.06	0.14	0.21	0.60	0.00	0.05	0.13	0.21
500	0.66	0.00	0.03	0.11	0.19	0.67	0.00	0.03	0.11	0.19	0.68	0.00	0.03	0.10	0.19
	True model: M_1 with $\theta_1 = 3$														
	Integral priors					Intrinsic priors					Akaike weights				
n	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
15	0.52	0.02	0.08	0.16	0.22	0.51	0.02	0.09	0.16	0.21	0.61	0.01	0.06	0.13	0.19
50	0.87	0.00	0.00	0.03	0.09	0.85	0.00	0.01	0.04	0.10	0.85	0.00	0.00	0.04	0.11
100	0.93	0.00	0.00	0.01	0.05	0.97	0.00	0.00	0.01	0.03	0.95	0.00	0.00	0.01	0.05
300	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
	True model: M_2 with $\theta_2 = 0.2$														
	Integral priors					Intrinsic priors					Akaike weights				
n	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
15	0.01	0.53	0.24	0.14	0.09	0.01	0.52	0.24	0.14	0.10	0.01	0.47	0.25	0.16	0.11
50	0.00	0.84	0.13	0.02	0.01	0.00	0.78	0.18	0.03	0.01	0.00	0.76	0.20	0.04	0.01
100	0.00	0.94	0.06	0.00	0.00	0.00	0.94	0.06	0.00	0.00	0.00	0.93	0.07	0.00	0.00

(*Continues*)

Table 5. *Continued*

	True model: M_2 with $\theta_2 = 0.5$														
	Integral priors					Intrinsic priors					Akaike weights				
n	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
15	0.13	0.29	0.22	0.19	0.18	0.11	0.34	0.21	0.18	0.16	0.12	0.31	0.21	0.18	0.17
50	0.04	0.46	0.23	0.15	0.12	0.04	0.51	0.22	0.13	0.10	0.04	0.49	0.23	0.14	0.10
100	0.01	0.66	0.20	0.08	0.05	0.00	0.64	0.22	0.09	0.05	0.00	0.65	0.20	0.09	0.05
300	0.00	0.89	0.10	0.01	0.00	0.00	0.89	0.10	0.01	0.00	0.00	0.86	0.13	0.01	0.00
	True model: M_3 with $\theta_2 = 0.2$														
	Integral priors					Intrinsic priors					Akaike weights				
n	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
15	0.00	0.17	0.33	0.28	0.22	0.01	0.20	0.31	0.27	0.22	0.01	0.16	0.30	0.28	0.24
50	0.00	0.12	0.55	0.24	0.08	0.00	0.09	0.60	0.23	0.08	0.00	0.07	0.59	0.25	0.09
100	0.00	0.03	0.74	0.21	0.03	0.00	0.02	0.81	0.16	0.01	0.00	0.01	0.80	0.17	0.02
	True model: M_3 with $\theta_2 = 0.5$														
	Integral priors					Intrinsic priors					Akaike weights				
n	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5	M_1	M_2	M_3	M_4	M_5
15	0.16	0.18	0.22	0.22	0.21	0.12	0.22	0.23	0.22	0.20	0.15	0.19	0.22	0.22	0.22
50	0.03	0.13	0.32	0.29	0.23	0.05	0.18	0.30	0.26	0.21	0.04	0.16	0.32	0.27	0.21
100	0.00	0.14	0.43	0.27	0.16	0.00	0.15	0.46	0.25	0.14	0.00	0.13	0.43	0.28	0.17
300	0.00	0.01	0.70	0.24	0.05	0.00	0.01	0.70	0.24	0.05	0.00	0.02	0.70	0.23	0.05

$$\pi_i(\theta_i) = \int \pi_i^N(\theta_i|z_i)p_i(z_i)\mathrm{d}z_i.$$

Therefore,

$$\pi_i(\theta_i, z_i|x) = \frac{f_i(x|\theta_i)\pi_i^N(\theta_i|z_i)p_i(z_i)}{m_i(x)}$$

is a completion of $\pi_i(\theta_i|x)$, and $\pi_i(\theta_i|z_i, x) = \pi_i^N(\theta_i|z_i, x)$. Then the Metropolis–Hastings algorithm with target $\pi_i(\theta_i, z_i|x)$ and proposal of the form

$$q_i(\theta_i', z_i'|\theta_i, z_i, x) = p_i(z_i')\rho_i(\theta_i'|\theta_i, z_i', x),$$

can be implemented provided

$$\frac{f_i(x|\theta_i')\pi_i^N(\theta_i'|z_i')\rho_i(\theta_i|\theta_i', z_i, x)}{f_i(x|\theta_i)\pi_i^N(\theta_i|z_i)\rho_i(\theta_i'|\theta_i, z_i', x)}. \tag{3}$$

can be computed. Note that if $\rho_i(\theta_i'|\theta_i, z_i', x)$ does not depend on z_i', then this Metropolis-Hasting algorithm becomes the pseudo-marginal Metropolis-Hastings algorithm of Andrieu & Roberts (2009) for the target $\pi_i(\theta_i|x)$, where the unbiased estimator of the target is $\pi_i(\theta_i, z_i|x)/p_i(z_i)$. If we choose

$$\rho_i(\theta_i'|\theta_i, z_i', x) = \pi_i^N(\theta_i'|z_i', x) = \frac{f_i(x|\theta_i')\pi_i^N(\theta_i'|z_i')}{m_i^N(z_i', x)/m_i^N(z_i')},$$

then (3) is given by

$$\frac{m_i^N(z_i', x)}{m_i^N(z_i')}\frac{m_i^N(z_i)}{m_i^N(z_i, x)}.$$

The aforementioned Metropolis-Hastings algorithm produces a Markov chain $(\tilde{\theta}_i^t, \tilde{z}_i^t)$ through the following transition. Given $\tilde{\theta}_i^t$ and $\tilde{z}_i^t$:

1. Take $z_i' = \check{z}_i^{t+1}$ and generate $\theta_i' \sim \rho_i(\theta_i'|\tilde{\theta}_i^t, \check{z}_i^{t+1}, x)$
2. Take $\tilde{\theta}_i^{t+1} = \theta_i'$ and $\tilde{z}_i^{t+1} = z_i'$ with probability

$$1 \wedge \frac{f_i(x|\theta_i')\pi_i^N(\theta_i'|z_i')\rho_i(\tilde{\theta}_i^t|\theta_i', \tilde{z}_i^t, x)}{f_i(x|\tilde{\theta}_i^t)\pi_i^N(\tilde{\theta}_i^t|\tilde{z}_i^t)\rho_i(\theta_i'|\tilde{\theta}_i^t, z_i', x)},$$

otherwise take $\tilde{\theta}_i^{t+1} = \tilde{\theta}_i^t$ and $\tilde{z}_i^{t+1} = \tilde{z}_i^t$.

Then the integral posterior and the integral prior can be estimated by

$$\hat{\pi_i}(\theta_i|x) = \frac{1}{\tilde{T}}\sum_{t=1}^{\tilde{T}} \pi_i^N(\theta_i|\tilde{z}_i^t, x),$$

and

$$\hat{\pi}_i(\theta_i) = \frac{1}{T}\sum_{t=1}^{T} \pi_i^N(\theta_i|z_i^t),$$

respectively.

This allows us to use Chib's (1995) approach to estimate the evidence by

$$\frac{\hat{\pi}_i(\theta_i^*) f_i(x|\theta_i^*)}{\hat{\pi}_i(\theta_i^*|x)}, \tag{4}$$

where θ_i^* is any plug-in value for θ_i.

We have applied this procedure to replicate the last row in Table 3 using the Monte Carlo estimator (1) and the Chib's method with the simulations from the integral posteriors with proposal $\rho_i(\theta_i'|\theta_i, z_i', x) = \pi_i^N(\theta_i'|z_i', x)$ in the Metropolis-Hastings algorithm. The number of simulations was set to $T = 10,000$ to approximate the integral priors and $\tilde{T} = 5000$ to approximate the integral posteriors. Using these approximations the evidence was estimated according to Equation 4, and the estimation was repeated 100 times. The resulting boxplots in Figure 8 illustrate the corresponding behaviour and the superiority of Chib's method.

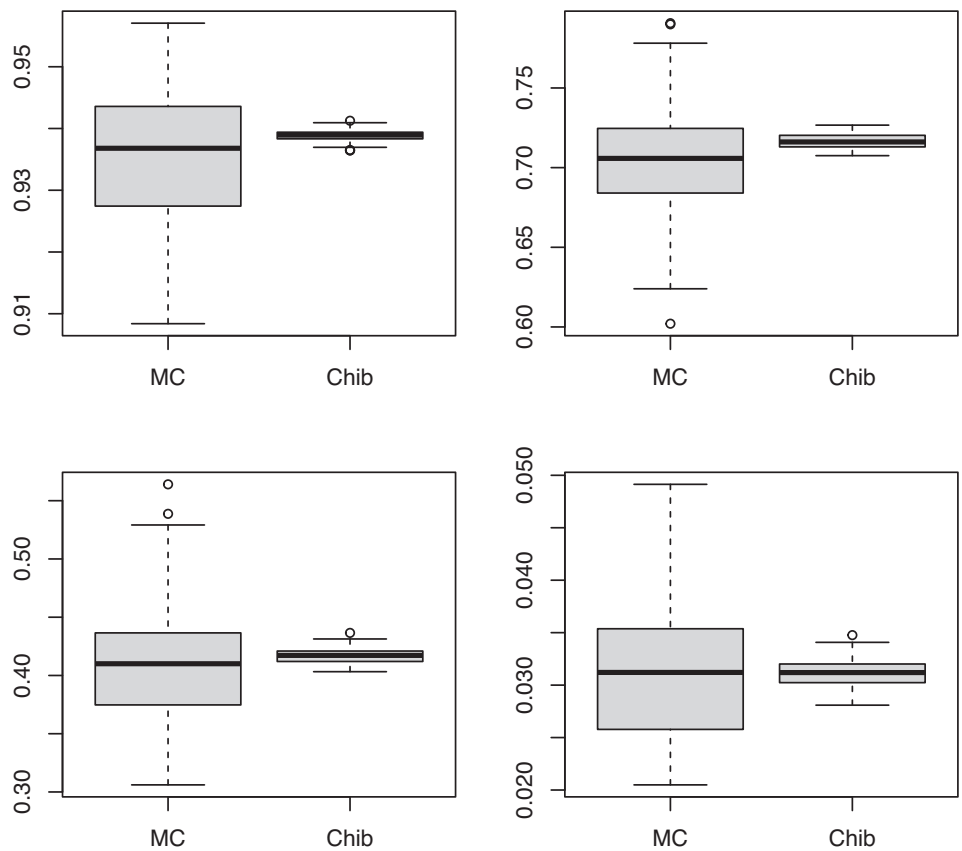

FIGURE 8. *Comparison of the estimates of the posterior probability of modelM_1 (last row in Table 3) using the estimator MC (1), and the Chib's method with the proposed Metropolis-Hastings algorithm to simulate from the integral posterior.*

8 Conclusions and Oncoming Research

We advocated in this paper a principled generalisation of the methodology of integral priors (Cano *et al.*, 2008) when two or more models are to be compared. Our approach proves to be quite automated and generic, both in the nested and in the non-nested cases, with the only requirement being the existence of training samples of finite size for all models under comparison. Although we have carried out some comparisons with more approximate solutions like intrinsic and fractional pseudo Bayes factors (Robert, 2001), further experiments could be conducted in that spirit. We recognise that the application of integral priors for complex models could involve certain degree of difficulty with respect to the concept of imaginary training samples, similar to the ones encountered with other automatic methods like expected posterior priors or intrinsic priors. On the other hand, usually integral priors have not an explicit form but must be obtained by simulation, which entails some computational limitations.

This methodology is an original proposal for the problem of choosing priors for model selection in the general case, rather than suggesting priors for specific cases. In this sense, our methodology falls within the class of approaches that produce default priors. Other methods, such as intrinsic priors, have been shown to work well for nested models, whereas for general non-nested models the problem has to be studied case-by-case. Our methodology is different in that it addresses all models simultaneously without privileging any one over another. However, we acknowledge that there is still much work to be done in the development of this theory, for example, improving computation time when the number of models is large, as well as the estimation of the evidence.

The contribution of our work, in relation to the prior developments by Cano & Salmerón (2013) and Cano *et al.* (2018), lies in having established conditions under which the integral priors are unique and proper prior distributions in such a way that the proposed methodology enables the application of the Metropolis-Hastings algorithm to simulate from the posterior distribution, thereby facilitating the use of general methods such as the harmonic mean estimator or Chib's method (Chib, 1995) to approximate the *evidence*, and consequently, the Bayes factor.

The examples we considered so far allow to approximate the Bayes factor with the Monte Carlo estimator, but extensions to more general settings are directly available when using evidence approximations (Friel & Wyse, 2012), like bridge sampling, inverse logistic regression or even harmonic means since some models (Marin & Robert, 2011) are associated with compact parameter spaces.

Occasionally researchers can express their knowledge about the parameters by establishing a range of possible values. In such cases, it would be possible to work with proper priors. If any of the models considered is equipped with such a proper prior, then it is not necessary to use artificial models with compact parametric spaces because the original model does not require imaginary training samples, and the prior would be the integral prior. Furthermore, we could use these ranges of values to construct the artificial models, in which case the integral priors could differ.

Acknowledgements

The research of D Salmerón was supported by the Fundación SéNeCa-Agencia de Ciencia y Tecnología de la Región de Murcia Program for Excellence in Scientific Research (project 20862/PI/18). CP Robert is partially funded by the European Union (ERC-2022-SYG-OCEAN-101071601) and by a PR[AI]RIE-PSAI chair from the Agence Nationale de la Recherche (ANR-23-IACL-0008).

References

Andrieu, C. & Roberts, G.O. (2009). The pseudo-marginal approach for efficient Monte Carlo computation. *Annals of Statistics*, **37**, 697–725.

Bayarri, M.J., Berger, J.O., Forte, A. & García-Donato, G. (2012). Criteria for Bayesian model choice with application to variable selection. *Annals of Statistics*, **40**, 1550–1577.

Bayarri, M.J. & García-Donato, G. (2008). Generalization of Jeffreys Divergence-Based Priors for Bayesian Hypothesis Testing. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **70**, 981–1003.

Berger, J.O. & Bernardo, J.M. 1992. On the development of reference priors (with discussion). In *Bayesian Statistics, 4*, Oxford University Press: London, pp. 35-60.

Berger, J.O., Bernardo, J.M. & Sun, D. (2024). *Objective Bayesian Inference*. World Scientific Pub Co Inc.

Berger, J.O. & Mortera, J. (1999). Default Bayes factors for nonnested hypothesis testing. *Journal of the American Statistical Association*, **94**(446), 542–554.

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

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

Cano, J.A., Iniesta, M. & Salmerón, D. (2018). Integral priors for Bayesian model selection: how they operate from simple to complex cases. *TEST*, **27**, 968–987.

Cano, J.A., Kessler, M. & Salmerón, D. (2007a). Integral priors for the one way random effects model. *Bayesian Analysis*, **2**(1), 59–68.

Cano, J.A., Kessler, M., Salmerón, D. et al. 2007b. A synopsis of integral priors for the one way random effects model. In *Bayesian Statistics*, Vol. **8**, Oxford University Press: Oxford, pp. 577-582.

Cano, J.A. & Salmerón, D. (2013). Integral priors and constrained imaginary training samples for nested and non-nested Bayesian model comparison. *Bayesian Analysis*, **8**(2), 361–380.

Cano, J.A. & Salmerón, D. (2016). A review of the developments on integral priors for Bayesian model selection. *BEIO*, **32**(2), 96–111.

Cano, J.A., Salmerón, D. & Robert, C.P. (2008). Integral equation solutions as prior distributions for Bayesian model selection. *TEST*, **17**(3), 493–504.

Chib, S. (1995). Marginal likelihoods from the Gibbs output. *Journal of the American Statistical Association*, **90**, 1313–1321.

Conigliani, C., Castro, J.I. & O'Hagan, A. (2000). Bayesian assessment of goodness of fit against nonparametric alternatives. *Canadian Journal of Statistics*, **28**, 327–342.

Consonni, G., Fouskakis, D., Liseo, B. & Ntzoufras, I. (2018). Prior distributions for objective Bayesian analysis. *Bayesian Analysis*, **13**, 627–679.

Fouskakis, D., Ntzoufras, I. & Draper, D. (2015). Power-Expected-Posterior priors for variable selection in Gaussian linear models. *Bayesian Analysis*, **10**, 75–107.

Friel, N. & Wyse, J. (2012). Estimating the evidence - a review. *Statistica Neerlandica*, **66**, 288–308.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2013). *Bayesian Data Analysis*. CRC press: Chapman & Hall.

Jeffreys, H. (1939). *Theory of Probability*, first edition. The Clarendon Press: Oxford.

Kass, R.E. & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, **90**, 928–934.

Marin, J.M. & Robert, C.P. 2011. Importance sampling methods for Bayesian discrimination between embedded models. In *Frontiers of Statistical Decision Making and Bayesian Analysis*, Eds. Chen, M.-H., Dey, D.K., Müller, P. & Sun, D., Springer: New York.

Moreno, E. (1997). Bayes factors for intrinsic and fractional priors in nested models. Bayesian Robustness. *Lecture Notes-Monograph Series*, **31**, 257–270.

Moreno, E. (2005). Objective Bayesian methods for one-sided testing. *TEST*, **14**, 181–198.

Moreno, E., Bertolino, F. & Racugno, W. (1998). An intrinsic limiting procedure for model selection and hypotheses testing. *Journal of the American Statistical Association*, **93**, 1451–1460.

Moreno, E., Martínez, C. & Vázquez-Polo, F. (2021). Objective Bayesian model choice for non-nested families: the case of the Poisson and the negative binomial. *TEST*, **30**, 255–273.

O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors. *TEST*, **6**, 101–118.

Pérez, J.M. & Berger, J.O. (2002). Expected posterior priors for model selection. *Biometrika*, **89**, 491–511.

Rissanen, J. (2012). *Optimal Estimation of Parameters*. Cambridge University Press.

Robert, C.P. (1995). Simulation of truncated normal variables. *Statistics and Computing*, **5**, 121–125.

Robert, C.P. (2001). *The Bayesian Choice*, second edition. Springer: New York.

Rousseau, J. et al. 2007. Approximating interval hypotheses: p-values and Bayes factors. In *Bayesian Statistics*, Vol. **8**, Oxford University Press: Oxford, pp. 427-462.
Salmerón, D., Cano, J.A. & Robert, C.P. (2015). Objective Bayesian hypothesis testing in binomial regression models with integral prior distributions. *Statistica Sinica*, **25**(3), 1009–1023.

JUAN ANTONIO CANO (1956-2018) was Professor in the Department of Statistics and Operations Research at the University of Murcia. Dr. Cano was a dear mentor and friend who contributed substantially to the theory of integral priors and was instrumental to the developments of this article.

[*Received June 2025; accepted January 2026*]