Bayesian Analysis (2007) **2**, Number 1, pp. 59–68

Integral priors for the one way random effects model

Juan Antonio Cano*, Mathieu Kessler† and Diego Salmerón‡

Abstract. The one way random effects model is analyzed from the Bayesian model selection perspective. From this point of view Bayes factors are the key tool to choose between two models. In order to produce objective Bayes factors objective priors should be assigned to each model.

However, these priors are usually improper provoking a calibration problem which precludes the comparison of the models. To solve this problem several derivations of automatically calibrated objective priors have been proposed among which we quote the intrinsic priors introduced in Berger and Pericchi (1996) and the integral priors introduced in Cano et al. (2006). Here, we focus on the use of integral priors which take advantage of MCMC techniques to produce most of the times unique Bayes factors. Some illustrations are provided.

Keywords: Bayesian model selection, Integral priors, Intrinsic priors, Random effects model, Recurrent Markov chains

1 Introduction

In model selection problems we consider two models M_i, $i = 1, 2$, where the data $\mathbf{x}$ are related to the parameter θ_i by a density $f_i(\mathbf{x}|\theta_i)$, the default (improper) priors are $\pi_i^N(\theta_i) = c_i h_i(\theta_i)$, $i = 1, 2$, and the resulting Bayes factor,

$$B_{21}^N(\mathbf{x}) = \frac{m_2^N(\mathbf{x})}{m_1^N(\mathbf{x})} = \frac{c_2 \int_{\Theta_2} f_2(\mathbf{x}|\theta_2)h_2(\theta_2)d\theta_2}{c_1 \int_{\Theta 1} f_1(\mathbf{x}|\theta_1)h_1(\theta_1)d\theta_1}$$

depends on the arbitrary ratio c_2/c_1. Therefore, we are left with two problems, that is, first the determination of the ratio c_2/c_1 is paramount, second the Bayes factor using $\pi_i^N(\theta_i)$ is not an actual Bayes factor and inference methods based on proper priors are preferable to those that are not, see Principle 1 in Berger and Pericchi (1996). An attempt for solving these problems (Berger and Pericchi (1996)), consists in using intrinsic priors $\pi_1^I$and π_2^I that are solutions to a system of two functional equations. Intrinsic priors provide a Bayes factor free of arbitrary constants but whether or not it is an actual Bayes factor or a limit of actual Bayes factors depends on the model (Berger and Pericchi, 1996). An additional difficulty when considering intrinsic priors is that they might be not unique, see Cano et al. (2004) and Cano et al. (2006).

*Universidad de Murcia, Murcia, Spain, mailto:jacano@um.es
†Universidad Politécnica de Cartagena, Cartagena, Spain, http://filemon.upct.es/~mathieu/
‡Consejería de Sanidad de Murcia, Murcia, Spain, mailto:salmeron@nsl.ethz.ch

 ba0005

On the other hand, in Cano et al. (2006) the so-called integral priors for model selection are proposed as solutions to the system of integral equations

$$\pi_1(\theta_1) = \int_{\mathcal{X}} \pi_1^N(\theta_1 \mid x) m_2(x) dx \tag{1}$$

and

$$\pi_2(\theta_2) = \int_{\mathcal{X}} \pi_2^N(\theta_2 \mid x) m_1(x) dx, \tag{2}$$

where x is an imaginary minimal training sample, see Berger and Pericchi (1996), and $m_i(x) = \int f_i(x|\theta_i)\pi_i(\theta_i)d\theta_i, i = 1, 2$. The method can be seen as a symmetrization of the equation that defines the expected posterior prior introduced in Pérez and Berger (2002). We emphasize that in this system both priors $\pi_i,\ i = 1, 2$, are the *incognita*.

These integral priors are well behaved, *i.e.* they provide a Bayes factor free of arbitrary constants which in fact is an actual Bayes factor or a limit of actual Bayes factors. Moreover, it turns out to be unique in many situations, and it can be shown that a sufficient condition for the uniqueness is that the Markov chain with transition density $Q(\theta_1' \mid \theta_1) = \int g(\theta_1, \theta_1', \theta_2, x, x')dxdx'd\theta_2$, where

$$g(\theta_1, \theta_1', \theta_2, x, x') = \pi_1^N(\theta_1' \mid x) f_2(x|\theta_2) \pi_2^N(\theta_2 \mid x') f_1(x'|\theta_1),$$

is recurrent.

In this paper, we consider the random effects model

$$M : y_{ij} = \mu + a_i + e_{ij}, i = 1, ..., k;\ j = 1, ..., n,$$

where the variables $e_{ij} \sim N(0, \sigma^2)$ and $a_i \sim N(0, \sigma_a^2)$, $i = 1, ..., k;\ j = 1, ..., n$, are independent. We are interested in the selection problem between models with parameters:

$$M_1 : \theta_1 = (\mu_1, \sigma_1, 0) \ \text{ and } \ M_2 : \theta_2 = (\mu_2, \sigma_2, \sigma_a).$$

The default priors we use to derive the integral priors in equations (1) and (2) are the reference priors $\pi_1^N(\theta_1) = c_1/\sigma_1$ and $\pi_2^N(\theta_2) = c_2\sigma_2^{-2}(1 + (\sigma_a/\sigma_2)^2)^{-3/2}$. Note that $\pi_1^N(\theta_1)$ is the reference prior for model M_1 and $\pi_2^N(\theta_2)$ is the reference prior for model M_2 for the ordered group $\{\sigma_a, (\sigma, \mu)\}$ when $n = 1$. We use the prior $\pi_2^N(\theta_2)$ for the sake of simplicity to keep the paper within a methodological level. Under these assumptions the sample densities for the two models are:

$$f_1(\mathbf{y} \mid \theta_1) = \prod_{i=1}^{k} N_n(\mathbf{y}_i \mid \mu_1 \mathbf{1}_n, \sigma_1^2 \mathbf{I}_n)$$

and

$$f_2(\mathbf{y}\mid\theta_2) = \prod_{i=1}^{k} N_n(\mathbf{y}_i \mid \mu_2\mathbf{1}_n, \sigma_2^2\mathbf{I}_n + \sigma_a^2\mathbf{J}_n),$$

where $\mathbf{y}_i = (y_{i1}, ..., y_{in})'$, $\mathbf{y} = (\mathbf{y}_1, ..., \mathbf{y}_k)'$, $\mathbf{1}_n = (1, ..., 1)'$, $\mathbf{I}_n$ is the identity matrix of dimension n and $\mathbf{J}_n$ the square matrix of dimension n with all the entries equal to one.

The rest of the paper is organized as follows. In section 2 the Bayes factor for the default priors is obtained. Integral and intrinsic priors for the one way random effects model are derived in section 3, where it is noticed that intrinsic priors are not unique. In section 4 we address the problem of the uniqueness of integral priors through the consideration of its associated Markov chain. In section 5 some illustrations are provided and finally, the conclusions are stated in section 6.

2 Bayes factors for the default priors

Let S be the total sum of squares, $\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\overline{y})^2$, decomposed as $S = S_1 + S_2$, where $S_1 = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\overline{y}_i)^2$ and $S_2 = \sum_{i=1}^{k} n(\overline{y}_i-\overline{y})^2$. The marginal $m_1^N(\mathbf{y})$ is obtained as

$$m_1^N(\mathbf{y}) \quad = \quad \int f_1(\mathbf{y}\mid\theta_1)\pi_1^N(\theta_1)d\theta_1 = \int \prod_{i=1}^{k} N_n(\mathbf{y}_i \mid \mu_1\mathbf{1}_n, \sigma_1^2\mathbf{I}_n)\frac{c_1}{\sigma_1}d\mu_1 d\sigma_1.$$

Applying theorem A.2.2 in Moreno and Torres (2004) it is straightforward to see that

$$m_1^N(\mathbf{y}) = c_1\frac{(S\pi)^{-\frac{nk-1}{2}}}{2}(nk)^{-\frac{1}{2}}\Gamma\left(\frac{nk-1}{2}\right).$$

On the other hand, $m_2^N(\mathbf{y})$ is given by

$$\int \prod_{i=1}^{k} N_n(\mathbf{y}_i \mid \mu_2\mathbf{1}_n, \sigma_2^2\mathbf{I}_n + \sigma_a^2\mathbf{J}_n)c_2\sigma_2^{-2}(1+(\sigma_a/\sigma_2)^2)^{-3/2}d\mu_2 d\sigma_2 d\sigma_a,$$

that, applying theorem A.2.3 in Moreno and Torres (2004) with $A_i = \sigma_2^2\mathbf{I}_n + \sigma_a^2\mathbf{J}_n$ and therefore $A_i^{-1} = \sigma_2^{-2}(\sigma_2^2+n\sigma_a^2)^{-1}\left[(\sigma_2^2+n\sigma_a^2)\mathbf{I}_n - \sigma_a^2\mathbf{J}_n\right]$, can be written as

$$m_2^N(\mathbf{y}) = c_2(2\pi)^{-\frac{nk-1}{2}}(nk)^{-\frac{1}{2}}\cdot$$

$$\cdot\int \sigma_2^{-(n-1)k}(\sigma_2^2+n\sigma_a^2)^{-\frac{k-1}{2}}\exp\left\{-\frac{S-\frac{n\sigma_a^2 S_2}{\sigma_2^2+n\sigma_a^2}}{2\sigma_2^2}\right\}\frac{(1+(\sigma_a/\sigma_2)^2)^{-3/2}}{\sigma_2^2}d\sigma_2 d\sigma_a.$$

Then using the change of variables $\sigma_a/\sigma_2 = u$ and $\sigma_2 = v$ we obtain

$$m_2^N(\mathbf{y}) = c_2(2\pi)^{-\frac{nk-1}{2}}(nk)^{-\frac{1}{2}}2^{\frac{nk-3}{2}}\Gamma\left(\frac{nk-1}{2}\right)S^{-\frac{nk-1}{2}}.$$

$$\cdot \int (1+nu^2)^{-\frac{k-1}{2}} \left(1 - \frac{nu^2}{1+nu^2}\frac{S_2}{S}\right)^{-\frac{nk-1}{2}} (1+u^2)^{-\frac{3}{2}} du.$$

The following expression is finally obtained for the Bayes factor $B_{21}^N(\mathbf{y})$

$$B_{21}^N(\mathbf{y}) = \frac{c_2}{c_1}\int_0^\infty (1+nu^2)^{-\frac{k-1}{2}} \left(1 - \frac{nu^2}{1+nu^2}\frac{S_2}{S}\right)^{-\frac{nk-1}{2}} (1+u^2)^{-\frac{3}{2}} du, \qquad (3)$$

which unfortunately as we previously stated depends on the arbitrary ratio c_2/c_1. To avoid this indeterminacy we will use integral and intrinsic priors instead of the original default priors.

3 Integral and intrinsic priors

To derive the integral and the intrinsic priors we first need an imaginary minimal training sample. According to the above expression for $m_1^N(\mathbf{y})$ and $m_2^N(\mathbf{y})$ we note that when $n = 1$ and $k = 2$, $m_1^N(\mathbf{y})$, $m_2^N(\mathbf{y}) > 0$, and therefore the vector $\mathbf{y}(l) = (y_{11}, y_{21})'$ is a minimal training sample; moreover, in this case $B_{21}^N(\mathbf{y}(l)) = 1$ and therefore $m_1^N(\mathbf{y}(l)) = m_2^N(\mathbf{y}(l))$. Consequently, see subsection 3.4 in Cano et al. (2006), $\{\pi_1^N, \pi_2^N\}$ are integral priors and they are well calibrated when $c_1 = c_2$.

Likewise, in Moreno et al. (1998) the following intrinsic priors $\{\pi_1^I, \pi_2^I\}$ are chosen

$$\begin{aligned} \pi_1^I(\theta_1) &= \pi_1^N(\theta_1) \\ \pi_2^I(\theta_2) &= \pi_2^N(\theta_2) E_{\mathbf{y}(l)|\theta_2}\{B_{12}^N(\mathbf{y}(l))\}, \end{aligned}$$

from where we deduce that $\{\pi_1^N, \pi_2^N\}$ are intrinsic priors and they are well calibrated when $c_1 = c_2$ too. Whether or not the integral and the intrinsic priors are unique is a matter of, at least, theoretical interest.

In the case where model M_1 is nested in model M_2, the system of functional equations mentioned in section 1 reduces to a single equation with two *incognita* and it is a well known result in the literature on intrinsic priors that they are not generally unique. Nevertheless, Moreno et al. (1998) provides a limiting procedure for choosing the above pair of sensible intrinsic priors that essentially consists in fixing $\pi_1^N(\theta_1)$ as the intrinsic prior for the simpler model obtaining $\pi_2^I(\theta_2)$ from the above mentioned functional equation as the intrinsic prior for model M_2. However, in Cano et al. (2006), it can be seen that integral priors are unique provided its associated Markov chain is recurrent.

4 Markov chain associated with the integral priors

The transition density of the associated Markov chain, $\theta_1 \rightarrow \theta_1'$, can be described in a simple way. The posterior distribution $\pi_1^N(\theta_1|\mathbf{y}(l))$ is obtained as

$$\pi_1^N(\sigma_1|\mathbf{y}(l)) \propto \sigma_1^{-2} \exp\left(-\frac{(y_{11}-y_{12})^2}{2\sigma_1^2}\right)$$

and

$$\pi_1^N(\mu_1|\sigma_1, \mathbf{y}(l)) = N(\mu_1|\overline{\mathbf{y}(l)}, \sigma_1^2/2),$$

where $\overline{\mathbf{y}(l)}$ is the average of the components of $\mathbf{y}(l)$.

To obtain the posterior distribution $\pi_2^N(\theta_2|\mathbf{y}(l))$ first note that

$$\pi_2^N(\mu_2|\sigma_a, \sigma_2, \mathbf{y}(l)) = N(\mu_2|\overline{\mathbf{y}(l)}, (\sigma_2^2 + \sigma_a^2)/2)$$

and

$$\pi_2^N(\sigma_2, \sigma_a|\mathbf{y}(l)) \propto \sigma_2^{-2}(1 + (\sigma_a/\sigma_2)^2)^{-3/2}(\sigma_a^2 + \sigma_2^2)^{-1/2} \exp\left(-\frac{(y_{11} - y_{12})^2}{4(\sigma_2^2 + \sigma_a^2)}\right),$$

then the random variables u, z and ε defined by the equations

$$\frac{\sigma_a}{\sigma_2} = u, \ \ \sigma_2^2 + \sigma_a^2 = \frac{(y_{11} - y_{12})^2}{4} z$$

and

$$\mu_2 = \overline{\mathbf{y}(l)} + \frac{1}{2}\varepsilon\sqrt{\sigma_2^2 + \sigma_a^2}$$

have densities $q_1(u) = (1 + u^2)^{-3/2}$, $q_2(z) \propto z^{-3/2}e^{-1/z}$ and $\varepsilon \sim N(0, 1)$. Therefore the simulation from $\pi_2^N(\theta_2|\mathbf{y}(l))$ can be done simulating u, z and ε and solving the above equations.

In summary, combining the transitions $\theta_1 \to \mathbf{y}(l)' \to \theta_2 \to \mathbf{y}(l) \to \theta_1'$ we obtain that

$$\mu_1' = \mu_1 + \sigma_1\alpha$$

and

$$\sigma_1' = \sigma_1\beta,$$

where

$$\beta = \frac{\sqrt{w}}{4} \mid \varepsilon_3 - \varepsilon_4 \mid \sqrt{z} \mid \xi_1 - \xi_2 \mid,$$

$$\alpha = \overline{\xi} + \frac{\varepsilon_2\sqrt{z}}{2\sqrt{2}} \mid \xi_1 - \xi_2 \mid + \frac{\varepsilon_3 + \varepsilon_4}{2}\sqrt{z} \mid \xi_1 - \xi_2 \mid /2 + \beta\varepsilon_1/\sqrt{2}$$

and $\xi_1, \xi_2, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4 \sim N(0, 1)$, $u \sim q_1(u) = (1 + u^2)^{-3/2}$, $z \sim q_2(z) \propto z^{-3/2}e^{-1/z}$ and $w \sim p(w) \propto w^{-3/2}e^{-1/w}$.

Regarding the autonomous chain (σ_n), $(\log \sigma_n)$ is a recurrent random walk since $E(\log \beta) = 0$ and $E((\log \beta)^2) < +\infty$. Although, for the whole chain (μ_n, σ_n), we have not been able to establish recurrence so far because the second order moments for α and β are not finite. However, under some assumptions commonly satisfied the recurrence of the chain can be established and then the integral priors are unique. This is done in the following proposition.

Proposition 1 *Let $\{\pi_1(\theta_1), \pi_2(\theta_2)\}$ be integral priors. Suppose that the integral prior $\pi_1(\theta_1) = \varphi(\sigma_1)$ does not depend on μ_1, then the integral priors are unique up to a multiplicative constant.*

Proof

It follows, see Cano et al. (2006), that $\pi_1(\theta_1)d\theta_1$ is an invariant σ-finite measure for the Markov chain

$$\mu_1' = \mu_1 + \sigma_1\alpha$$
$$\sigma_1' = \sigma_1\beta.$$

If $p_2(\beta)p_1(\alpha|\beta)$ denotes the density function of (α, β) then the transition density $Q(\theta_1'|\theta_1)$ is

$$\frac{1}{\sigma_1^2}p_2\left(\frac{\sigma_1'}{\sigma_1}\right)p_1\left(\frac{\mu_1'-\mu_1}{\sigma_1}\Big|\frac{\sigma_1'}{\sigma_1}\right)$$

and it follows from the invariance property that

$$\varphi(\sigma_1) = \int \frac{1}{\sigma_1}p_2\left(\frac{\sigma_1'}{\sigma_1}\right)p_1\left(r\Big|\frac{\sigma_1'}{\sigma_1}\right)\varphi(\sigma_1')drd\sigma_1' = \int \frac{1}{\sigma_1}p_2\left(\frac{\sigma_1'}{\sigma_1}\right)\varphi(\sigma_1')d\sigma_1'.$$

Therefore $\varphi(\sigma_1)d\sigma_1$ is an invariant σ-finite measure for the recurrent Markov chain $\sigma_1' = \sigma_1\beta$ meaning that $\varphi(\sigma_1)$, and therefore $\pi_1(\theta_1)$, have to be proportional to $1/\sigma_1$, and the proposition is proved.

Remark 1. *Note that if the integral prior $\pi_1(\theta_1)$ can be written as $\pi_1(\theta_1) = \varphi_1(\mu_1|\sigma_1)\varphi_2(\sigma_1)$ with*

$$\int \varphi_1(\mu_1|\sigma_1)d\mu_1 = 1, \, \forall\sigma_1 > 0,$$

then, from the invariance property it follows that

$$\varphi_2(\sigma_1) = \int \frac{1}{\sigma_1}p_2\left(\frac{\sigma_1'}{\sigma_1}\right)\varphi_2(\sigma_1')d\sigma_1'$$

and again $\varphi_2(\sigma_1)$, and therefore $\pi_1(\theta_1)$, have to be proportional to $1/\sigma_1$ and the integral priors are unique up to a multiplicative constant.

5 Illustrations

In the context of variance components problems, two popular data sets can be found in Box and Tiao (1973), pages 246 and 247, respectively, where a parameter estimation approach is used. We compute the Bayes factors for the two sets of data, using equation (3), with $c_1 = c_2$.

The first set of data in Box and Tiao (1973) concerns batch to batch variation in yields of dyestuff. The data arise from a balanced experiment where the total product yield was determined for 5 samples from each of 6 randomly chosen batches of raw material. The object of the study was to determine the relative importance of between batch variation versus variation due to sampling and analytic errors. To illustrate the difficulty of the analysis and the need for objective procedures, we have explored the sensitivity of the Bayes factor to the choice of hyperparameters of proper priors. More concretely, assume we choose the following conventional proper priors:

$$\begin{aligned} M_1: &\quad (\mu_1, \sigma_1^2) \sim \mathcal{N}(\bar{y}, \sigma_0^2) \otimes \mathcal{IG}(\alpha, \beta), \\ M_2: &\quad (\mu_2, \sigma_2^2, \sigma_a^2) \sim \mathcal{N}(\bar{y}, \sigma_0^2) \otimes \mathcal{IG}(\alpha, \beta) \otimes \mathcal{IG}(\alpha_a, \beta_a), \end{aligned}$$

where $\mathcal{IG}(\alpha, \beta)$ denotes the Inverse Gamma distribution with mean $\beta/(\alpha-1)$ and variance $\beta^2/((\alpha-1)^2(\alpha-2))$. Fixing $\alpha = 20$, $\alpha_a = 13$, $\beta_a = 20000$ and varying β, we find the following numerical values for the associated Bayes factors when $\sigma_0 \to +\infty$:

$\beta \times 10^{-3}$	25	30	35	40	45	50
B_{21}	2.99	1.90	1.32	0.98	0.77	0.63

It turns out the Bayes factor is very sensitive to the choice of the hyperparameters and may lead to opposite conclusions.

The Bayes factor when the integral priors described in this paper are chosen is found to be $B_{21} = 10.0179$, which means evidence in favor of M_2.

We also computed for this data set the Bayes factor B_{21}^{WG} when the priors suggested in Westfall and Gönen (1996) are chosen. An expression similar to (3) is obtained and the numerical value is found to be $B_{21}^{WG} = 11.9$.

The second set of data are simulated data with $\mu_2 = 5$, $\sigma_2 = 4$, $\sigma_a = 2$, $k = 6$ and $n = 5$. For these data the resulting Bayes factor is $B_{21} = 0.3671$, which means evidence in favor of model M_1 despite the fact that the data were generated from model M_2. Notice that, for these data, we find $B_{21}^{WG} = 0.17$. An explanation for the misbehavior of these Bayes factors is that the standard deviation of the error σ_2 is twice as big as the standard deviation σ_a of the random effects, which, together to the small number of groups, makes difficult the detection of the random effects. A graphical exploration of the data set, see Figure 1, confirms that even if the groups centers are different the intragroup variation is too big to allow detection. We also have tested the normality (Shapiro-Wilk test) of these data and the resulting *p-value* is 0.735. All these results are in agreement with the results in Box and Tiao (1973), where it is found that the posterior density for σ_a has its mode at the origin and is monotonically decreasing.

6 Conclusions

The methodology of integral priors introduced in Cano et al. (2006) has been applied to solve the one way random effects model from a Bayesian perspective. The main conclusion is that the default priors $\left\{\pi_1^N, \pi_2^N\right\}$ are both integral priors and intrinsic priors and they are well "calibrated" when $c_1 = c_2$. This allows to use them to compute the Bayes factor which has been done with two sets of data.

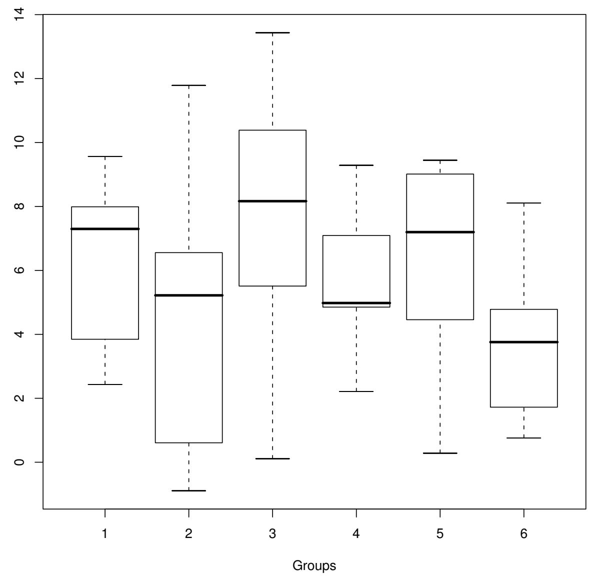

Figure 1: Boxplot of the simulated data in Box and Tiao (1973), p 247: the intra-group variability overwhelms the known inter-group variability.

The Markov chain associated when we use this methodology provides some insight on how integral and intrinsic (in this case) priors work out. Finally, we note that intrinsic priors are not unique while under some commonly satisfied assumptions the integral priors are unique. However, the question of whether or not the integral priors for this problem are unique is still an open problem.

References

Berger, J. O. and Pericchi, L. R. (1996). "The intrinsic Bayes factor for model selection and prediction." *J. Amer. Statist. Assoc.*, 91: 109–122. 59, 60

Box, G. E. P. and Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont. 64, 65

Cano, J. A., Kessler, M., and Moreno, E. (2004). "On intrinsic priors for nonnested models." *Test*, 13(2): 445–463. 59

Cano, J. A., Salmerón, D., and Robert, C. P. (2006). "Integral equation solutions as prior distributions for Bayesian model selection." *Accepted for publication in Test*. 59, 60, 62, 64, 65

Moreno, E., Bertolino, F., and Racugno, W. (1998). "An intrinsic limiting procedure for model selection and hypotheses testing." *J. Amer. Statist. Assoc.*, 93(444): 1451–1460. 62

Moreno, E. a. and Torres, F. (2004). "Resultados asociados a la normal multivariante." *Universidad de Granada*, (Technical Report). 61

Pérez, J. M. and Berger, J. O. (2002). "Expected-posterior prior distributions for model selection." *Biometrika*, 89(3): 491–511. 60

Westfall, P. H. and Gönen, M. (1996). "Asymptotic properties of ANOVA Bayes factors." *Comm. Statist. Theory Methods*, 25(12): 3101–3123. 65