Available on line at www.rac.es/racsam
Statistics and Operations Research

REVISTA DE LA REAL ACADEMIA DE CIENCIAS
EXACTAS, FISICAS Y NATURALES.
SERIE A: MATEMATICAS
Madrid (España / Spain)

RACSAM **104** (1), 2010, 57–67. DOI:10.5052/RACSAM.2010.06

Consistency of objective Bayes factors for nonnested linear models and increasing model dimension

F. Javier Girón, Elías Moreno, George Casella and M. L. Martínez

Abstract Casella et al. [2, (2009)] proved that, under very general conditions, for normal linear models the Bayes factor for a wide class of prior distributions, including the intrinsic priors, is consistent when the number of parameters does not grow with the sample size n. The special attention paid to the intrinsic priors is due to the fact that they are nonsubjective priors, and thus accessible priors for complex models.

The case where the number of parameters of nested models grows as $O(n^\alpha)$ for $\alpha \leq 1$ was considered in Moreno et al. [13, (2010)], in which it was proved that the Bayes factor for intrinsic priors is consistent for the case where both models are of order $O(n^\alpha)$ for $\alpha < 1$, and for $\alpha = 1$ is consistent except for a small set of alternative models. The small set of models for which consistency does not hold was characterized in terms of a pseudo-distance between models.

The goal of the present article is to extend the above results to the case where the linear models are nonnested. As the comparison of nonnested models calls for a method of encompassing, for proving consistency we use encompassing from below in this paper.

Consistencia de factores de Bayes objetivos para modelos lineales anidados cuando la dimensión de los modelos crece

Resumen. En Casella et al. [2, (2009)] se demostró que, bajo condiciones muy generales, el factor de Bayes para modelos lineales normales y para una amplia clase de distribuciones a priori, que incluía a las a priori intrínsecas, es consistente cuando el número de parámetros no crece cuando lo hace el tamaño muestral n. Se prestó especial atención a las distribuciones a priori intrínsecas debido a que son distribuciones a priori no subjetivas y, por consiguiente, se pueden aplicar a modelos complejos.

El caso en que el número de parámetros de los modelos anidados crece del orden de $O(n^\alpha)$ para $\alpha \leq 1$ se ha considerado en Moreno et al. [13, (2010)], en el que se demuestra que el factor de Bayes para distribuciones intrínsecas es consistente para el caso en que ambos modelos son de orden $O(n^\alpha)$ para $\alpha < 1$ y, para el caso $\alpha = 1$, también es consistente excepto para un conjunto pequeño de modelos alternativos. Este conjunto, para el cual la consistencia no se da, se caracterizó en términos de una pseudo distancia entre modelos.

Submitted by David Ríos
Received: December 28, 2009. *Accepted*: February 3, 2010
Keywords: Asymptotics, BIC, intrinsic priors, linear models, multiplicity of parameters, rate of growth
Mathematics Subject Classifications: 62F05, 62J15

El objetivo del presente artículo es extender los resultados anteriores al caso en que los modelos lineales no estén anidados. Como la comparación de modelos no anidados requiere algún método de *abarcamiento*, para demostrar la consistencia en el artículo usamos generalmente el método de *abarcamiento por abajo*.

1 Introduction

A widely used objective quasi-Bayesian procedure for comparing pairs of nested models is the well-known BIC (Bayesian Information Criterion) introduced by Schwarz [15, (1978)]. This procedure is based on an approximation to the Bayes factor derived from the application of the Laplace approximation to the marginals of the data under the competing models. In spite of its simplicity, it is an accurate tool for comparing normal linear models of low dimension when the available sample has a large size.

A recent objective Bayesian model selector competitor to the BIC is the Bayes factor for intrinsic priors. The objective intrinsic prior distributions were introduced by Berger and Pericchi [1, (1996)], and formalized in Moreno [9, (1997)] and Moreno et al. [10, (1998)]. For normal linear models the derivation of the intrinsic priors and the corresponding Bayes factor was given in Casella and Moreno [4, (2006)] and Girón, Martínez and Moreno [5, (2006)]. Later on, in Casella et al. [2, (2009)] it was proved that for normal linear models the Bayes factor is a consistent model selector for a wide class of prior distributions including the intrinsic priors. The Bayes factor for intrinsic priors was also proven to be a superior model selector than the BIC for finite sample sizes, and equivalent to the BIC asymptotically. Throughout that paper it was assumed that the dimension of the models did not grow with the sample size.

The problem of comparing nonnested models has also been considered in the literature; for instance, in Girón, Martínez and Moreno [5, (2006)] two different encompassing criteria, named *encompassing from above* and *encompassing from below*, were analyzed. The idea behind the encompassing from below was to compare each of the two nonnested models with a smaller fixed model which is encompassed in both, and the idea behind the encompassing from above was to compare each of them with a larger common model that encompasses both. A comparative analysis of both procedures of encompassing in the class of normal linear regression model was given in Moreno and Girón [11, (2008)] and the conclusion was that these encompassing methods behave extremely well for the important variable selection problem in which nonnested models have to be compared.

On the other hand, change-point detection in time series —see Smith [14, (1975)], Moreno, Casella, and García-Ferrer [12, (2005)], and Girón, Moreno and Casella [6, (2007)]— or clustering (Hartigan [7, (1990)], Casella, Girón and Moreno [3, (2010)]) are problems such that the dimension of the underlying models grows as the sample size grows, and this suggests to inquiring about the consistency of the BIC and the Bayes factor for intrinsic priors for this case. This question has been addressed in a recent paper by Moreno, Girón and Casella [13, (2010)] where it was proved that, under very general conditions, for comparing nested normal linear models both the BIC and the Bayes factor for intrinsic priors are consistent when the number of parameters is of order $O(n^\alpha)$ with $\alpha < 1$; when the dimension of the alternative model is of order $O(n)$ the BIC is inconsistent and the Bayes factor for intrinsic priors is almost consistent, where almost consistency is understood as consistency except for a small set of alternative models. The small set of alternative models was also characterized in terms of a pseudo-distance from the alternative to the null model.

In this paper we will extend some of the results obtained for the case of nested models when the dimension of the models grows with the sample size to the case of nonnested models. We note that in most of the model selection problems, comparison among nested and nonnested models naturally appear, so that to address the consistency issue for nonnested models is of the utmost importance.

In Section 2 the consistency of the Bayes factor for intrinsic priors when the models are nonnested is proved for $\alpha < 1$. Likewise, the BIC is also proved to be consistent for nonnested models under the same conditions. Almost consistency of the Bayes factor for intrinsic priors is proved for two models when one is of order $O(n^a)$ for $a < 1$ and the other of other $O(n)$, and the set of models around the smallest one for which consistency does not hold is characterized. In this case, however, the BIC is proved to be consistent under the smallest model but inconsistent under the largest model. We use an encompassing from below procedure, although it seems that encompassing from above produces equivalent results in some cases. Section 3 provides a short concluding discussion.

2 Consistency for nonnested linear models

In this section we introduce some notation and some previous results needed to prove the main theorems in the paper. We begin with a statement of the problem of model comparison for nested linear normal models.

Let $\mathbf{y} = (y_1, \ldots, y_n)'$ be a vector of independent responses, $\mathbf{X}_p$ a design matrix of dimension $n \times p$, where p is the number of deterministic explanatory variables, and let $\mathbf{X}_i$ denote a submatrix of $\mathbf{X}_p$ whose dimensions are $n \times i$. We want to compare the reduced sampling model $M_i : N_n(\mathbf{y}|\mathbf{X}_i\alpha_i, \sigma_i^2\,\mathbf{I}_n)$, and the full model $M_p : N_n(\mathbf{y}|\mathbf{X}_p\beta_p, \sigma_p^2\,\mathbf{I}_n)$, where $N_n(\mathbf{y}|\mu, \mathbf{\Sigma})$ denotes the multivariate normal distribution of dimension n with mean vector μ and covariance matrix $\mathbf{\Sigma}$. We assume that the regression parameter vectors $\alpha_i = (\alpha_1, \ldots, \alpha_i)',\ \beta_p = (\beta_1, \ldots, \beta_p)'$, and the variance errors σ_i^2, σ_p^2, are unknown. Note that with these assumptions the reduced model M_i is nested in the full model M_p. The Bayes factor for comparing these models —see, for instance, Moreno, Girón and Casella [13, (2010)]—, turns out to be

$$B_{pi}(\mathbf{y}) = \frac{2}{\pi}\int_0^{\pi/2}\left(1 + \frac{n}{(p+1)\sin^2\varphi}\right)^{(n-p)/2}\left(1 + \frac{n\mathcal{B}_{ip}}{(p+1)\sin^2\varphi}\right)^{-(n-i)/2} \mathrm{d}\varphi.$$

where the statistic $\mathcal{B}_{ip}$ is

$$\mathcal{B}_{ip} = \frac{\mathrm{RSS}_p}{\mathrm{RSS}_i} = \frac{\mathbf{y}'(\mathbf{I}_n - \mathbf{H}_p)\mathbf{y}}{\mathbf{y}'(\mathbf{I}_n - \mathbf{H}_i)\mathbf{y}},$$

and $\mathbf{H}_j = \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'$, for $j = i, p$, the hat matrix.

One important quantity in the study of consistency is the directed “distance” from the regression model M_p to M_i, which is an extension of the one first introduced in Casella et al. [2, (2009)] to account for models for which the number of parameters and the sample size increase to infinity, which is defined for a given sample size n as

$$\delta_{pi} = \frac{1}{\sigma_p^2}\,\beta_p'\,\frac{\mathbf{X}_p'(\mathbf{I}_n - \mathbf{H}_i)\mathbf{X}_p}{n}\,\beta_p.$$

Note that this pseudo-distance is defined for every pair of models not necessarily nested, and it is not symmetric. Some useful properties of δ_{pi}, which will be used in the sequel, are the following:

a) The distance from any model M_i to itself is always 0.

b) If M_i is nested in M_j, then $\delta_{ij} = 0$.

c) If model M_i is nested in M_j, then $\delta_{ki} \geq \delta_{kj}$ for any model M_k.

In the study of the asymptotic behavior of the Bayes factor $B_{pi}(\mathbf{y})$, we will have to consider the limit of the distance as n grows to infinity. This limit, assuming it exists, will be denoted by

$$\delta_{pi}^* = \lim_{n\to\infty}\frac{1}{\sigma_p^2}\,\beta_p'\,\frac{\mathbf{X}_p'(\mathbf{I}_n - \mathbf{H}_i)\mathbf{X}_p}{n}\,\beta_p.$$

In what follows $\lim_{n\to\infty}[M]Z_n$ will denote the limit in probability of the random sequence $\{Z_n; n \geq 1\}$ under the assumption that we are sampling from model M. This model M will have a fixed parameter sequence. Further, we will need to use the doubly noncentral beta distribution with parameters $\upsilon_1/2, \upsilon_2/2$ and noncentrality parameters λ_1, λ_2, which will be denoted by $\mathrm{Be}(\upsilon_1/2, \upsilon_2/2; \lambda_1, \lambda_2)$ (Johnson, Kotz and Balakrishnan [8, (1995), pp. 502]).

Next we state Lemma 1, which is the basic tool used in this section for deriving the theorems comparing nonnested models. The proof is similar to that of Lemma 1 of Moreno et al. [13, (2010)], and therefore is omitted.

Lemma 1

1) *If model M_i is nested in model M_p, and M_t is the true model, then the sampling distribution of the statistic $\mathcal{B}_{ip}$ under M_t is the noncentral beta distribution*

$$\mathcal{B}_{ip} \sim \mathrm{Be}\left(\frac{n-p}{2}, \frac{p-i}{2}; n\delta_{tp}, n(\delta_{ti}-\delta_{tp})\right).$$

2) *Let $\{X_n, n \geq 1\}$ be a sequence of random variables such that*

$$X_n \sim \mathrm{Be}\left(\frac{n-p}{2}, \frac{p-i}{2}; n\delta_{tp}, n(\delta_{ti}-\delta_{tp})\right), \qquad n \geq 1.$$

If i and p vary with n as $i = O(n^a)$ and $p = O(n^b)$, where $0 \leq a \leq b \leq 1$, then when sampling from M_t we have,

(i) *if $a < b = 1$ we have that*

$$\lim_{n\to\infty} [M_t]\ X_n = \frac{1+\delta^*_{tp} - 1/r}{1+\delta^*_{ti}},$$

where r is a positive constant such that $r = \lim_{p\to\infty} n/p > 1$.

(ii) *If $a = b = 1$ we have that*

$$\lim_{n\to\infty} [M_t]\ X_n = \frac{1+\delta^*_{tp} - 1/r}{1+\delta^*_{ti} - 1/s},$$

where the positive constants r and s are such that $r = \lim\limits_{p\to\infty} n/p > 1$ and $s = \lim\limits_{p\to\infty} n/i > 1$.

(iii) *If $b < 1$, then we have*

$$\lim_{n\to\infty} [M_t]\ X_n = \frac{1+\delta^*_{tp}}{1+\delta^*_{ti}}.$$

The second tool we need is an asymptotic approximation to the Bayes factor for intrinsic priors. We derive this asymptotic approximation for the case where $i = O(n^a)$ and $p = O(n^b)$ and $0 \leq a \leq b < 1$, using a similar approach as the one used in the proof of Theorem 2 in Moreno et al. [13, (2010)]. Note that when i and p are finite, the first two terms of the approximation is precisely the BIC criterion up to a finite multiplicative constant.

Lemma 2 *If model M_i is nested in model M_p, and $b < 1$, then for large values of n the Bayes factor $B_{pi}(\mathbf{y})$ can be approximated by*

$$B_{pi}(\mathbf{y}) \approx \exp\left\{\left(\frac{i-p}{2}\right)\log\frac{n}{p+1} - \left(\frac{n-i}{2}\right)\log\mathcal{B}_{ip} + \left(\frac{p+1}{2}\right)\left(1-\frac{1}{\mathcal{B}_{ip}}\right)\right\}.$$

PROOF. Factoring the integrand of the Bayes factor $B_{pi}(\mathbf{y})$ as

$$\left(\frac{n}{(p+1)\sin^2\varphi}\right)^{\frac{n-p}{2}}\left(\frac{n\mathcal{B}_{ip}}{(p+1)\sin^2\varphi}\right)^{\frac{i-n}{2}}\left(1+\frac{(p+1)\sin^2}{n}\right)^{\frac{n-p}{2}}\left(1+\frac{(p+1)\sin^2}{n\mathcal{B}_{ip}}\right)^{\frac{i-n}{2}}$$

and, taking into account that

$$\frac{(p+1)}{n} \qquad \text{and} \qquad \frac{(p+1)}{n\mathcal{B}_{ip}}$$

go to zero as n tends to infinity, the third and fourth factors can be approximated by exponentials, and the Bayes factor is approximately equal to

$$\frac{2}{\pi}\left(\frac{n}{p+1}\right)^{\frac{i-p}{2}}\mathcal{B}_{ip}^{-(\frac{n-i}{2})}\int_0^{\pi/2}\sin\varphi^{p-i}\exp\left[\left(\frac{p+1}{2n}\right)\left(n-p+\frac{i-n}{\mathcal{B}_{ip}}\right)\sin^2\varphi\right]\mathrm{d}\varphi.$$

But the term in brackets is approximated by

$$\frac{p+1}{2}\left(1-\frac{1}{\mathcal{B}_{ip}}\right)\sin^2\varphi;$$

so that the new approximation is

$$\frac{2}{\pi}\left(\frac{n}{p+1}\right)^{\frac{i-p}{2}}\mathcal{B}_{ip}^{-(\frac{n-i}{2})}\int_0^{\pi/2}\sin\varphi^{p-i}\exp\left[\frac{p+1}{2}\left(1-\frac{1}{\mathcal{B}_{ip}}\right)\sin^2\varphi\right]\mathrm{d}\varphi.$$

By the mean value theorem, the integral above is equal to $\pi/2$ times the value of the integrand at some value of $\varphi(p,i,\mathcal{B}_{ip})$. It can be shown that, except for values near $\mathcal{B}_{ip}=1$ or $\mathcal{B}_{ip}=0$ where the integral goes to infinity, and for large values of p an i, the values of $\varphi(p,i,\mathcal{B}_{ip})$ converge to $\pi/2$ and consequently the values of $\sin\varphi(p,i,\mathcal{B}_{ip})$ approach to 1, and thus the integral can be approximated by

$$\frac{\pi}{2}\exp\left\{\frac{p+1}{2}\left(1-\frac{1}{\mathcal{B}_{ip}}\right)\right\}.$$

Substituting the integral by this expression, we finally get the desired approximation for the Bayes factor of this Lemma 2. ■

Under the conditions of Lemma 1 and 2, we provide a further approximation to the Bayes factor for large values of n, when sampling from the true model, as follows.

Lemma 3 *If model M_i is nested in model M_p, $b<1$, and M_t is the true model, then for large values of n and under the sampling model M_t, the Bayes factor can be approximated by*

$$B_{pi}(\mathbf{y})\approx\exp\left\{\left(\frac{i-p}{2}\right)\log\frac{n}{p+1}\right\}\left(\frac{1+\delta_{ti}}{1+\delta_{tp}}\right)^{n/2}.$$

PROOF. First we note that when $b<1$, the third term in the expression of the approximation in Lemma 2 is of order $O(n^b)$, while the first and second are of order $O(n\log n)$ and $O(n)$, respectively; therefore, the leading terms of the approximation are

$$\begin{aligned}B_{pi}(\mathbf{y})&\approx\exp\left\{\left(\frac{i-p}{2}\right)\log\frac{n}{p+1}-\left(\frac{n-i}{2}\right)\log\mathcal{B}_{ip}\right\}\\&\approx\exp\left\{\left(\frac{i-p}{2}\right)\log\frac{n}{p+1}\right\}\left(\frac{1+\delta_{ti}}{1+\delta_{tp}}\right)^{n/2}.\end{aligned}$$

■

As a curiosity we note that from this latter approximation it is very simple to give a simple direct proof of the consistency of the Bayes factor for intrinsic priors in the nested case, which was stated as Corollary 4 in Moreno et al. [13, (2010)].

2.1 Consistency when $\dim(\mathbf{M}_i) = \mathbf{O}(\mathbf{n}^a)$ for $a < 1$ and $\dim(\mathbf{M}_j) = \mathbf{O}(\mathbf{n}^b)$ for $b < 1$

We are now in a position to prove consistency of the Bayes factor for intrinsic priors for nonnested models by encompassing them from below. Let us consider an arbitrary pair of nonnested models M_i and M_j and let M_0 denote a model nested in both M_i and M_j. Depending on the model selection problem the encompassing model M_0 is chosen in a different form. For instance, in the variable selection the intercept only model is used as the encompasing model M_0; in the changepoint problem M_0 is the no changepoint model, and in clustering the model M_0 is the one cluster model.

Let us define the from below Bayes factor for comparing M_j and M_i, conditional on M_0, as

$$B'_{ji}(\mathbf{y}|M_0) = \frac{B_{j0}(\mathbf{y})}{B_{i0}(\mathbf{y})}.$$

Next Theorem 1 states the consistency when either of the two models is the true one.

Theorem 1 *Let M_0 be a model nested in both nonnested models M_i and M_j, where $i = O(n^a)$ and $j = O(n^b)$ and $a, b < 1$. If $\delta^*_{ij} > 0$ and $\delta^*_{ji} > 0$, then, the from below Bayes factor is consistent under both models.*

PROOF. If the true model is M_i, then from Lemma 3 and the fact that $\delta^*_{ii} = 0$, discarding lower order terms we have

$$B'_{ji}(\mathbf{y}|M_0) = \frac{B_{j0}(\mathbf{y})}{B_{i0}(\mathbf{y})} \approx \exp\left[\left(\frac{i-j}{2}\right)\log n\right] \cdot \left(\frac{1+\delta_{i0}}{1+\delta_{ij}}\right)^{n/2} \Big/ \left(\frac{1+\delta_{i0}}{1+\delta_{ii}}\right)^{n/2}$$
$$\approx \exp\left[\left(\frac{i-j}{2}\right)\log n\right](1+\delta_{ij})^{-n/2}.$$

As $\delta^*_{ij} > 0$, we obtain that $\lim_{n\to\infty}[M_i]\ B'_{ji}(\mathbf{y}) = 0$, thus proving consistency under model M_i. In the same way, by symmetry, we prove that if $\delta^*_{ji} > 0$, then $\lim_{n\to\infty}[M_j]\ B'_{ji}(\mathbf{y}) = \infty$. This completes the proof of the theorem. ■

Remark 1 *Note that although the value of the from below Bayes factor $B'_{ji}(\mathbf{y}|M_0)$ depends on the chosen model M_0, this theorem demonstrates that from the consistency point of view the selection of the model M_0 is irrelevant, provided it is nested in both M_i and M_j.*

Remark 2 *We know that when M_i is nested in M_j, then it follows that $\delta_{ij} = 0$ for all n. Then, when sampling from the reduced model M_i, the approximation of the from below intrinsic Bayes factor is*

$$B'_{ji}(\mathbf{y}) \approx \exp\left[\left(\frac{i-j}{2}\right)\log n\right]$$

and, as $i < j$, we get consistency under the null. Consistency under the alternative is guaranteed when $\delta^*_{ji} > 0$. The case where the distance from one model to the other, say $\delta^*_{ji} = 0$, in the nonnested case is much more involved and will be dealt with elsewhere.

Remark 3 *In a similar fashion, it can be proved that if we encompass both models M_i and M_j* from above *by a model M_F the dimension of which is of order $O(n^b)$ for $b < 1$, the resulting Bayes factor is also consistent under the same assumptions as those of Theorem 1, whatever the encompassing model M_F should be.*

From this theorem we can prove that the BIC is consistent for some nonnested models.

Corollary 1 *Under the condition of theorem 1, the BIC is consistent under both models provided that* $\delta^*_{ij} > 0$ *and* $\delta^*_{ji} > 0$.

PROOF. The BIC or Schwarz [15, (1978)] criterion for linear models is defined as

$$S_{ji}(\mathbf{y}) = \exp\left\{ \left(\frac{i-j}{2}\right) \log n - \frac{n}{2} \log \mathcal{B}_{ij} \right\},$$

where $\mathcal{B}_{ij} = \mathrm{RSS}_j \,/\, \mathrm{RSS}_i \in (0, \infty)$. Writing,

$$\mathcal{B}_{ij} = \frac{\mathrm{RSS}_j \,/\, \mathrm{RSS}_0}{\mathrm{RSS}_i \,/\, \mathrm{RSS}_0}$$

where RSS_0 is the residual sum of squares of any encompassing model M_0, the Schwarz approximation $S_{ji}(\mathbf{y})$ is written as

$$S_{ji}(\mathbf{y}) = \exp\left\{ \left(\frac{i-j}{2}\right) \log n - \frac{n}{2} \log \frac{\mathcal{B}_{0j}}{\mathcal{B}_{0i}} \right\},$$

and from Lemma 3 we get that under M_i we have

$$S_{ji}(\mathbf{y}) \approx \exp\left[\left(\frac{i-j}{2}\right) \log n \right] (1 + \delta_{ij})^{-n/2},$$

which is exactly the same asymptotic approximation of $B'_{ji}(\mathbf{y}|M_0)$. By symmetry the same argument applies to M_j, and the corollary follows suit. ■

2.2 Consistency when the $\dim(\mathbf{M}_i) = \mathbf{O}(\mathbf{n}^a)$ for $a < 1$ and $\dim(\mathbf{M}_j) = \mathbf{O}(\mathbf{n})$

Theorem 1 studies consistency for the case where the model M_i is of order $O(n^a)$, $a < 1$, and M_j of order $O(n^b)$, $b < 1$. In this section we explore consistency when the dimension of one of the nonnested model is of order $O(n)$. We note that now the dimension of the encompassing model M_0 should be of order $O(n^\alpha)$ for $\alpha \leq a$. The next lemma will be needed for the proof of Theorem 2.

Lemma 4 *Suppose that* M_i *is nested in* M_p *and* $i = O(n^a)$, $p = O(n)$, *with* $a < 1$. *Let* M_t *denote the true model. Then, the asymptotic approximation of the Bayes factor* $B_{pi}(\mathbf{y})$ *under* M_t *is*

$$B_{pi}(\mathbf{y}) \approx \left[(1+r)^{\frac{r-1}{r}} \left(1 + \frac{r(1+\delta_{tp}) - 1}{1+\delta_{ti}} \right)^{-1} \right]^{n/2}.$$

PROOF. The proof follows by using the same derivation of the asymptotic approximation of the Bayes factor in Moreno et al. [13, (2010)] with the difference of using part (ii) of Lemma 1 instead of part (i). ■

The next theorem provides a sufficient condition for the consistency of the Bayes factor in the nonnested case.

Theorem 2 *Let* M_0 *be any model nested in both* M_i *and* M_j *such that* $i = O(n^a)$ *with* $a < 1$, *and* $j = O(n)$, *and* $r = \lim_{n\to\infty} n/j > 1$.

i) *If* $\delta^*_{ij} > 0$, *then the from below Bayes factor is consistent under* M_i.

ii) *If* $\delta^*_{j0} > \delta(r)$, *where*

$$\delta(r) = \frac{r-1}{(r+1)^{(r-1)/r} - 1} - 1,$$

and

$$\delta^*_{ji} \in \left(\frac{r + \delta^*_{j0}}{(1+r)^{\frac{r-1}{r}}} - 1, \delta^*_{j0} \right],$$

then, the from below Bayes factor is consistent under M_j.

PROOF. We recall that the from below Bayes factor $B'_{ji}(\mathbf{y})$ for comparing M_i and M_j is defined as $B'_{ji}(\mathbf{y}) = B_{j0}(\mathbf{y})/B_{i0}(\mathbf{y})$. Because model M_0 is of order $O(n^\alpha)$ for $\alpha < 1$, the Bayes factor $B_{i0}(\mathbf{y})$ under M_i is approximated by

$$B_{i0}(\mathbf{y}) \approx \exp\left\{ -\frac{i}{2} \log \frac{n}{i+1} \right\} \left(\frac{1+\delta_{i0}}{1+\delta_{ii}} \right)^{n/2},$$

and discarding the first factor —the lower order term— and recalling that $\delta_{ii} = 0$, it is finally approximated by

$$B_{i0}(\mathbf{y}) \approx (1+\delta_{i0})^{n/2}.$$

On the other hand, by Lemma 4 it follows that $B_{j0}(\mathbf{y})$ under M_i is approximated by

$$B_{j0}(\mathbf{y}) \approx \left[(1+r)^{\frac{r-1}{r}} \left(1 + \frac{r(1+\delta_{ij}) - 1}{1+\delta_{i0}} \right)^{-1} \right]^{n/2}.$$

Thus, the approximation to the from below Bayes factor turns out to be

$$B'_{ji}(\mathbf{y}) \approx \left[(1+r)^{\frac{r-1}{r}} \left(r(1+\delta_{ij}) + \delta_{i0} \right)^{-1} \right]^{n/2},$$

but, from property c) of the pesudo-distance, we have that $\delta_{i0} \geq \delta_{ij}$. Therefore, the term in brackets of the from below Bayes factor is smaller than

$$\frac{(1+r)^{\frac{r-1}{r}}}{r + (1+r)\delta_{ij}}.$$

But this function of the two variables r and δ_{ij}, subject to the constraints $r \geq 1$ and $\delta_{ij} \geq 0$, is a positive function and bounded from above by 1. As the maximum is attained at $r = 1$ and $\delta_{ij} = 0$, and since $\delta^*_{ij} > 0$, the term in brackets of the from below Bayes factor is strictly smaller than 1, so that the from below Bayes factor tends to 0 as n grows, and, consequently, consistency under M_i is proved.

Consistency under model M_j, subject to the stated constraints, is proved as follows. Under M_j, the Bayes factor $B_{i0}(\mathbf{y})$ is approximated by

$$B_{i0}(\mathbf{y}) \approx \left(\frac{1+\delta_{j0}}{1+\delta_{ji}} \right)^{n/2}.$$

On the other hand, the Bayes factor $B_{j0}(\mathbf{y})$ under M_j is approximated by

$$B_{j0}(\mathbf{y}) \approx \left[(1+r)^{\frac{r-1}{r}} \left(1 + \frac{r(1+\delta_{jj}) - 1}{1+\delta_{j0}} \right)^{-1} \right]^{n/2},$$

but, as $\delta_{jj} = 0$, after some simplifications, the from below Bayes factor can be written as

$$B'_{ji}(\mathbf{y}) \approx \left[(1+r)^{\frac{r-1}{r}} \left(\frac{1+\delta_{ji}}{r+\delta_{j0}} \right) \right]^{n/2}.$$

For the from below Bayes factor to be consistent, it is sufficient that the term in brackets be strictly larger than 1 when n tends to infinity. This is equivalent to

$$(1+r)^{\frac{r-1}{r}}(1+\delta_{ji}) > r + \delta_{j0},$$

which implies that

$$\delta_{ji} > \frac{r+\delta_{j0}}{(1+r)^{\frac{r-1}{r}}} - 1.$$

On the other hand, from property c) of the distance, we have $\delta_{ji} \leq \delta_{j0}$.

In order for the interval where the distance δ_{ji} should lie

$$\delta_{ji} \in \left(\frac{r+\delta_{j0}}{(1+r)^{\frac{r-1}{r}}} - 1, \delta_{j0} \right]$$

to be nonempty, a necessary and sufficient condition is that $\delta_{j0} > \delta(r)$ for all $r \geq 1$. Thus, taking limits, under the conditions stated in the theorem, the from below Bayes factor tends to infinity, and the theorem is proved.

Remark 4 *In the proof of the second part of Theorem 2 it is interesting to note that the consistency under M_j depends on the model M_0 through the distance δ_{j0} only. Further, it is easy to prove that, under the conditions of Lemma 1 part (ii), the limit of the distance δ_{ji} of the large model M_j to the small model M_i, when n tends to infinity, also satisfies the condition $\delta^*_{ji} > \delta(r)$. On the other hand, part ii) of Theorem 2 suggests that the encompasing model M_0 should be chosen as small as possible in order to maximize the distance δ^*_{j0}.*

This lower bound on the distances δ^*_{j0} and δ^*_{ji} is exactly the same that the one obtained in Theorem 2 of Moreno et al. [13, (2010)]. *Moreover, assuming that $M_0 = M_i$, this latter theorem, which is valid when M_i is nested in M_j, can be now obtained as a particular case of our more general Theorem 2.*

The behavior of the BIC under the conditions of Theorem 2 is shown to be such that, although it is consistent under the smallest model, it is never consistent under the largest model.

Theorem 3 *Under the conditions of Theorem 2, the BIC is always consistent under M_i, and it is always inconsistent under M_j*

PROOF. The BIC

$$S_{ji}(\mathbf{y}) = \exp\left\{ \left(\frac{i-j}{2}\right)\log n - \frac{n}{2}\log \mathcal{B}_{ij} \right\},$$

under the stated conditions, implies that $j \approx n/r$, and that $\mathcal{B}_{ij}$ converges in probability to a finite positive number under both models, so that the leading term of the Schwarz approximation is the first one

$$S_{ji}(\mathbf{y}) \approx \exp\left\{ -\left(\frac{n}{2r}\right)\log n \right\},$$

because the first term in the exponent of the BIC is a non-random quantity of order $O(n \log n)$, while the second stochastic term is of order $O(n)$. As n goes to infinity, $S_{ji}(\mathbf{y})$ always goes to 0. This implies that the BIC is always consistent under M_i and always inconsistent under M_j.

Remark 5 *The preceding argument is valid as far as δ^*_{ji} is finite, a condition which is assumed in section 2. If δ^*_{ji} is allowed to be infinite, then $\mathcal{B}_{ij}$ tends to 0 in probability, and we could obtain consistency under the alternative depending on the rate of divergence of the limiting pseudo-distance δ^*_{ji}. This means that if the distance of the largest model to the smallest one is enormous, then the BIC could choose the former model with probability 1.*

3 Discussion

We have extended previous results on the consistency of the Bayes factor for intrisic priors obtained for the case of nested normal linear models to the case where the models are nonnested. We have considered several scenarios depending on the order of magnitude of the dimensions of the models involved using the from below way of encompassing. In some cases consistency of the from below Bayes factor for any encompassing model M_0 can be obtained; in others, consistency may depend on the distance from the larger model to the encompassing model.

The case where both models are of order $O(n)$ still remains an open problem, for in this case there might be many candidates to be the encompassing model, which should have dimension of order $O(n^{\alpha})$ for $\alpha \leq 1$. Yet, we believe that to explore this case the basic tools are those presented in Lemmas 1 and 2.

On the other hand, Lemma 2 provides an approximation to the Bayes factor for intrinsic priors for nested model that it is worthwhile to be analyzed, since it gives surprising numerical accurate results that need to be explored and will be addressed somewhere.

Acknowledgement. Elías Moreno is supported by Ministerio de Ciencia y Tecnología, Grant SEJ-65200 and Junta de Andalucía Grant SEJ-02814.

George Casella is supported by National Science Foundation Grants DMS-04-05543, DMS-0631632 and SES-0631588.

References

[1] Berger, J. O. and Pericchi, L. R., (1996). The intrinsic Bayes factor for model selection and prediction, *J. Amer. Statist. Assoc.*, **91**, 109–122. DOI: 10.2307/2291387

[2] Casella, G.; Girón, F. J.; Martínez, M. L. and Moreno, E., (2009). Consistency of Bayesian procedures for variable selection, *Ann. Statist.*, **37**, 1207–1228. DOI: 10.1214/08-AOS606

[3] Casella, G.; Girón, F. J. and Moreno, E., (2010). Cluster Analysis through model selection (submitted).

[4] Casella, G. and Moreno E., (2006). Objective Bayesian variable selection. *J. Amer. Statist. Assoc.*, **101**, 157–167. DOI: 10.1198/016214505000000646

[5] Girón, F. J.; Martínez, M. L. and Moreno, E., (2006). An objective Bayesian procedure for variable selection in regression. In N. Balakrishnan, E. Castillo and J. M. Sarabia, eds., *Advances in Distribution Theory, Order Statistics, and Inference*. 389–404, Birkhauser: Boston.

[6] Girón, F. J.; Moreno, E. and Casella, G., (2007). A Bayesian analysis of multiple change points in linear models (with discussion). In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West (Eds.), *Bayesian Statistics*, **8**, 227–252, Oxford University Press.

[7] Hartigan, J. A., (1990). Partition Models, *Comm. Statist. Theory and Methods*, **19**, 8, 2745–2756. DOI: 10.1080/03610929008830345

[8] Johnson, N. L.; Kotz, S. and Balakrishnan, N., (1995). *Continuous univariate distributions*, Vol. II, second edition, Wiley: New York.

[9] Moreno, E., (1997). Bayes factor for intrinsic and fractional priors in nested models: Bayesian robustness, In L_1-*Statistical Procedures and Related Topic*, Institute of Mathematical Statistics: Hayward, CA, 257–270. DOI: 10.1214/lnms/1215454142

[10] Moreno, E.; Bertolino, F. and Racugno, W., (1998). An intrinsic limiting procedure for model selection and hypothesis testing, *J. Amer. Statist. Assoc.*, **93**, 1451–1460. DOI: 10.2307/2670059

[11] Moreno, E. and Girón, F. J., (2008). Comparison of Bayesian objective procedures for variable selection in linear regression. *Test*, **17**, 472–492. DOI: 10.1007/s11749-006-0039-1

[12] MORENO, E.; CASELLA, G. AND GARCÍA-FERRER, A., (2005). An objective Bayesian analysis of the change point problem, *Stochastic Environmental Research Risk Assessment*, **19**, 191–204. DOI: 10.1007/s00477-004-0224-2

[13] MORENO, E.; GIRÓN, F. J. AND CASELLA, G., (2010). Consistency of objective Bayes tests as the model dimension grows. *Ann. Statist*, (to appear).

[14] SMITH, A. F. M., (1975). A Bayesian approach to inference about change-point in a sequence of random variables, *Biometrika*, **62**, 407–416.DOI: 10.1093/biomet/62.2.407

[15] SCHWARZ, G., (1978). Estimating the dimension of a model, *Ann. Statist.*, **6**, 461–464. DOI: 10.1214/aos/1176344136

F. Javier Girón
Professor,
Department of Statistics,
University of Málaga,
29071 Málaga
Spain.
fj_giron@uma.es

Elías Moreno
Professor,
Department of Statistics,
University of Granada,
18071, Granada,
Spain.
emoreno@ugr.es

George Casella
Distinguished Professor,
Department of Statistics,
University of Florida,
Gainesville, FL 32611
USA.
casella@stat.ufl.edu

M. L. Martínez
Associate Professor,
Department of Statistics,
University of Málaga,
29071 Málaga
Spain.
mlmartinez@uma.es