RAC$\mathbb{SAM}$
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **97 (**1), 2003, pp. 53–61
Estadística e Investigación Operativa / Statistics and Operations Research

Intrinsic priors for hypothesis testing in normal regression models

Elías Moreno, F. Javier Girón and Francisco Torres

Abstract. Testing that some regression coefficients are equal to zero is an important problem in many applications. Homoscedasticity is not necessarily a realistic condition in this setting and, as a consequence, no frequentist test there exist. Approximate tests have been proposed.

In this paper a Bayesian analysis of this problem is carried out, from a default Bayesian model choice perspective. Explicit expressions for intrinsic priors are provided, and it is shown that the corresponding Bayes factor is computed with the help of very simple numerical computations.

Distribuciones a priori intrínsecas para modelos de regresión normales

Resumen. En muchas aplicaciones es frecuente enfrentarse con el problema de contrastar si algunos coeficientes de regresión son nulos. Dicho problema se resuelve, bajo el punto de vista frecuentista, imponiendo la hipótesis de homocedasticidad. Sin embargo esta suposición no es asumible en general, proporcionándose en tales casos tests aproximados.

En este artículo se realiza un análisis bayesiano de este problema a partir de la perspectiva bayesiana de selección de modelos. Se obtienen expresiones explícitas para las distribuciones a priori intrínsecas y se comprueba que los factores Bayes asociados se reducen a expresiones fácilmente calculables con métodos numéricos.

1. Introduction

Suppose that the observable random variable Y follows a normal regression model with k covariates $x_1, \ldots, x_k$. That is, the density of a sample $(y_i, x_{i1}, \ldots, x_{ik})$, $i = 1, \ldots, n$, is $N_n(\mathbf{y}|\mathbf{X}\alpha, \sigma^2\mathbf{I}_n)$, where $\mathbf{y} = (y_1, \ldots, y_n)^t$, α is the column vector made up with the intercept α_0 and the regression coefficients $\alpha_1, \ldots, \alpha_k$ of the covariates $x_1, \ldots, x_k$, $\mathbf{I}_n$ is the identity matrix of size $n \times n$, and $\mathbf{X}$ is the $n \times (k+1)$ design matrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix}.$$

The nuisance parameter σ is unknown. The data is denoted as $\mathbf{D} = (\mathbf{y}, \mathbf{X})$.

The main motivation in considering this model comes from the analysis of matched pairs in Girón *et al.* [5]. In the matched pairs problem the vector $\mathbf{y}$ is typically a difference between a posttreatment and

Presentado por Sixto Ríos García.
Recibido: 17 de Julio de 2002. *Aceptado*: 9 de Octubre de 2002.
Palabras clave / Keywords: Bayes factor, Intrinsic priors, Model choice.
Mathematics Subject Classifications: 62F15, 62J05

the pretreatment measure drawn from each individual of a population of n individuals and each of them receives a dose of a treatment, possibly multidimensional, that here might be represented by a subvector of the covariates vector. Furthermore, the population is assumed to be a nonhomogeneous group, so that the treatment effect may depend on the dose levels of the treatment and on some characteristics of the patients measured by a set of values of a certain set of covariates. In that paper was shown that testing the null hypothesis that a subvector α_0 of α of length k_0 has all its components equal to zero plays a central role in the matched pairs problem.

Let us consider the partition of α as $\alpha = (\alpha_0^t, \alpha_1^t)^t$ and let $\mathbf{X} = (\mathbf{X}_0|\mathbf{X}_1)$ be the corresponding partition of the columns of the design matrix $\mathbf{X}$. Thus, $\mathbf{X}_0$ is a $n \times k_0$ matrix and $\mathbf{X}_1$ a $n \times k_1$ matrix, where $k_1 = k + 1 - k_0$. Then, the above testing problem can be formulated as

$$H_0 : \alpha_0 = \mathbf{0} \ \textit{versus} \ H_1 : \alpha_0 \neq \mathbf{0}. \tag{1}$$

We remark that when $\alpha_0 = \mathbf{0}$ the meaning of the remaining coefficients in α changes and, thus, we will denote them as γ_1. It is also clear that the nuisance parameter σ under the null and under the alternative is not necessarily the same. Therefore, homoscedasticity of the sampling models under H_0 and H_1 should not be imposed. Hence, under the null the sampling model for $\mathbf{y}$ is a multivariate normal distribution with mean vector $\mathbf{X}_1\gamma_1$ and covariance matrix $\sigma_0^2\mathbf{I}_n$, that is $\mathbf{y} \backsim N_n(\mathbf{y}|\mathbf{X}_1\gamma_1, \sigma_0^2\mathbf{I}_n)$, and under the alternative the sampling model for $\mathbf{y}$ is a $N_n(\mathbf{y}|\mathbf{X}\alpha, \sigma_1^2\mathbf{I}_n)$. As a consequence no frequentist test does exist, although some approximations are available (see, for instance, Chow (1960)).

Assuming vague prior information on (γ_1, σ_0) and (α, σ_1) the testing problem 1 can be formulated as a model selection problem between the two conventional Bayesian models

$$M_0 : N_n(\mathbf{y}|\mathbf{X}_1\gamma_1, \sigma_0^2\mathbf{I}_n), \ \ \pi_0^N(\gamma_1, \sigma_0) = c_0/\sigma_0^2,$$

and

$$M_1 : N_n(\mathbf{y}|\mathbf{X}\alpha, \sigma_1^2\mathbf{I}_n), \ \ \pi_1^N(\alpha, \sigma_1) = c_1/\sigma_1^2,$$

where c_0 and c_1 are arbitrary positive constant that cannot be determined because the priors are improper. For the usual $0-1$ loss function, model M_0 is chosen if $P(M_0|\mathbf{D}) > P(M_1|\mathbf{D})$. The model posterior probabilities are computed assuming the conventional default prior $P(M_0) = P(M_1) = 1/2$ on the set of models $\{M_0, M_1\}$. Then, we can write

$$P(M_0|\mathbf{D}) = \frac{1}{1 + B_{10}^N(\mathbf{D})},$$

where $B_{10}^N(\mathbf{D})$ represents the Bayes factor for comparing model M_1 and M_0, and is given by

$$B_{10}^N(\mathbf{D}) = \frac{\int N_n(\mathbf{y}|\mathbf{X}\alpha, \sigma_1^2\mathbf{I}_n)\pi_1^N(\alpha, \sigma_1)d\alpha d\sigma_1}{\int N_n(\mathbf{y}|\mathbf{X}_1\gamma_1, \sigma_0^2\mathbf{I}_n)\pi_0^N(\gamma_1, \sigma_0)d\gamma_1 d\sigma_0}.$$

The latter expression shows that the usual reference priors cannot be used for model choice because they leave the Bayes factor defined up to the arbitrary multiplicative constant c_1/c_0. The goal of this paper is to derive sensible priors for computing a well-defined Bayes factor for comparing models M_1 and M_0. In section 2 it is shown that the intrinsic methodology can be applied to deriving intrinsic priors from the conventional $\pi_0^N(\gamma_1, \sigma_0)$ and $\pi_1^N(\alpha, \sigma_1)$ considered above. In section 3 the Bayes factor for intrinsic priors is found and we observe that it requires very few computational efforts. Some conclusions are stated in the final section 4.

2. Intrinsic priors

The intrinsic prior methodology was introduced by Berger and Pericchi [1] to overcome the difficulty arising with the conventional prior in model selection problems. The difficulty arises because the conventional priors for the candidate models are typically improper and then the Bayes factor is defined up to a multiplicative constant, as occurs in testing regression coefficients considered above. Further developments of the method were given by Moreno, Bertolino and Racugno [7].

Justifications for the use of intrinsic priors for model selection have been given by Berger and Pericchi ([1], [2], [3], [4]), Moreno [6], and Moreno Bertolino and Racugno [7]. The method has proven to provide sensible prior for a wide variety of model selection problems involving nested models; see, for instance, Berger and Pericchi ([1], [4]), Girón et al. [5], Moreno, Bertolino and Racugno ([7], [8], [9]), Moreno and Liseo [10]), Moreno, Torres and Casella [11].

We note that the sampling model $N_n(\mathbf{y}|\mathbf{X}_1\gamma_1, \sigma_0^2\mathbf{I}_n)$ is nested in $N_n(\mathbf{y}|\mathbf{X}\alpha, \sigma_1^2\mathbf{I}_n)$. This allows for the application of the intrinsic methodology. Following Moreno, Bertolino and Racugno [7], the intrinsic priors for model M_0 and M_1 can be derived as follows. Let us consider a theoretical training sample of minimal size for the full model, say $\{(z_j, x_{1j}, \ldots, x_{kj}),\ j = 1, \ldots, k+1\}$. Hence, we have $\mathbf{z} \backsim N_{k+1}(\mathbf{z}|\mathbf{Z}\alpha, \sigma_1^2\mathbf{I}_{k+1})$, where $\mathbf{z} = (z_1, \ldots, z_{k+1})^t$, and

$$\mathbf{Z} = \begin{pmatrix} x_{11} & \ldots & x_{1k} \\ x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots \\ x_{k+1,1} & \ldots & x_{k+1,k} \end{pmatrix}.$$

Then, the intrinsic prior for α, σ_1 conditional on an arbitrary but fixed point γ_1, σ_0 is given by

$$\pi^I(\alpha, \sigma_1|\gamma_1, \sigma_0) = \pi_1^N(\alpha, \sigma_1) E_{\mathbf{z}|\alpha,\sigma_1} B_{01}^N(\mathbf{z}), \tag{2}$$

where the expectation is taken with respect to the density $N_{k+1}(\mathbf{z}|\mathbf{Z}\alpha, \sigma_1^2\mathbf{I}_{k+1})$.

By construction $\pi^I(\alpha, \sigma_1|\gamma_1, \sigma_0)$ is a probability density for any point γ_1, σ_0. The unconditional intrinsic prior for α, σ_1 is obtained by integrating out γ_1 and σ_0 with respect to the conventional prior $\pi_0^N(\gamma_1, \sigma_0)$, that is

$$\pi^I(\alpha, \sigma_1) = \int \pi^I(\alpha, \sigma_1|\gamma_1, \sigma_0)\pi_0^N(\gamma_1, \sigma_0) d\gamma_1 d\sigma_0.$$

The pair $(\pi_0^N(\gamma_1, \sigma_0), \pi^I(\alpha, \sigma_1))$ is called the intrinsic priors for comparing M_0 and M_1 and although they are improper, (i) they are well calibrated since both depend on the same arbitrary constant c_0, and (ii) they are a well established limit of proper priors.

In what follows, the moment generating function of a p-variate normal distribution $N_p(\mathbf{y}|\theta, \mathbf{\Sigma})$ will be denoted as $M_{\mathbf{y},\theta,\mathbf{\Sigma}}(t)$.

Lemma 1 *Let $\mathbf{X}$ be a $p \times q$ matrix of rank q ($q \leq p$) and $\mathbf{\Sigma}$ a $p \times p$ positive definite matrix. Then*

$$\int_{\mathbb{R}^q} N_p(\mathbf{y}|\theta, \mathbf{\Sigma}) d\theta = \frac{\exp\left(-\frac{1}{2}\mathbf{y}^t \left[\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{\Sigma}^{-1}\right]\mathbf{y}\right)}{(2\pi)^{\frac{p-q}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}|\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{X}|^{\frac{1}{2}}}.$$

PROOF. Expanding the quadratic form in $N_p(\mathbf{y}|\theta, \mathbf{\Sigma})$, it is immediate to see that

$$\begin{aligned} \int_{\mathbb{R}^q} N_p(\mathbf{y}|\theta, \mathbf{\Sigma}) d\theta &= \frac{\exp\left(-\frac{1}{2}\mathbf{y}^t\mathbf{\Sigma}^{-1}\mathbf{y}\right)|\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{X}|^{-\frac{1}{2}}}{(2\pi)^{\frac{p-q}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}} \int_{\mathbb{R}^q} \exp\left(\mathbf{y}^t\mathbf{\Sigma}^{-1}\mathbf{X}\theta\right) N_q(\theta|0, (\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}) d\theta \\ &= \frac{\exp\left(-\frac{1}{2}\mathbf{y}^t\mathbf{\Sigma}^{-1}\mathbf{y}\right)|\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{X}|^{-\frac{1}{2}}}{(2\pi)^{\frac{p-q}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}} M_{\theta,0,(\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}}(\mathbf{X}^t\mathbf{\Sigma}^{-1}\mathbf{y}), \end{aligned}$$

and the result follows. ■

Corollary 1 *Let $\mathbf{X}$ be a $p \times q$ matrix of rank q ($q \leq p$), h a natural number and $\mathbf{P} = \mathbf{I}_p - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$. Then*

$$\int_{\mathbb{R}^q}\int_{\mathbb{R}^+} \frac{1}{\sigma^h} N_p(\mathbf{y}|\mathbf{X}\theta, \sigma^2\mathbf{I}_p) d\theta d\sigma = \frac{2^{\frac{h-3}{2}}\Gamma\left(\frac{p+h-q-1}{2}\right)}{\pi^{\frac{p-q}{2}}|\mathbf{X}^t\mathbf{X}|^{\frac{1}{2}}(\mathbf{y}^t\mathbf{P}\mathbf{y})^{\frac{p+h-q-1}{2}}}.$$

PROOF. From lemma 1 the double integral becomes

$$\int_{\mathbb{R}^+} \frac{\exp\left(-\frac{1}{2\sigma^2}\mathbf{y}^t\mathbf{P}\mathbf{y}\right)}{(2\pi)^{\frac{p-q}{2}}(\sigma^2)^{\frac{p+h-q}{2}}|\mathbf{X}^t\mathbf{X}|^{\frac{1}{2}}} d\sigma.$$

Using the transformation $t = \frac{\mathbf{y}^t\mathbf{P}\mathbf{y}}{2\sigma^2}$, the preceding integral becomes

$$\frac{2^{\frac{h-3}{2}}}{\pi^{\frac{p-q}{2}}|\mathbf{X}^t\mathbf{X}|^{\frac{1}{2}}(\mathbf{y}^t\mathbf{P}\mathbf{y})^{\frac{p+h-q-1}{2}}}\int_{\mathbb{R}^+} t^{\frac{p+h-q-1}{2}-1}e^{-t}dt = \frac{2^{\frac{h-3}{2}}\Gamma\left[\frac{p+h-q-1}{2}\right]}{\pi^{\frac{p-q}{2}}|\mathbf{X}^t\mathbf{X}|^{\frac{1}{2}}(\mathbf{y}^t\mathbf{P}\mathbf{y})^{\frac{p+h-q-1}{2}}}.$$

This proves corollary 1. ■

Lemma 2 *Let $\mathbf{K}$ be a $p \times p$ symmetric matrix, and for $i = 1, 2$, let $\mathbf{X}_i$ be $p \times q_i$ matrices of rank q_i, with $q_i \leq p$, θ_i vectors of length q_i and $\mathbf{A}_i$ symmetric positive definite matrices of dimensions $p \times p$, then*

$$\int_{\mathbb{R}^p}\left(\mathbf{y}^t\mathbf{K}\mathbf{y}\prod_{i=1}^{2} N_p(\mathbf{y}|\mathbf{X}_i\theta_i, \mathbf{A}_i)\right) d\mathbf{y} = \left[tr(\mathbf{E}^{-1}\mathbf{K}) + \mathbf{D}^t\mathbf{E}^{-1}\mathbf{K}\mathbf{E}^{-1}\mathbf{D}\right] N_p\left[\mathbf{X}_2\theta_2|\mathbf{X}_1\theta_1, \mathbf{A}_1 + \mathbf{A}_2\right],$$

where $\mathbf{D} = \sum_{i=1}^{2}\mathbf{A}_i^{-1}\mathbf{X}_i\theta_i$ and $\mathbf{E} = \mathbf{A}_1^{-1} + \mathbf{A}_2^{-1}$.

PROOF. Since $\mathbf{A}_1$ and $\mathbf{A}_2$ are positive definite matrices, then $\mathbf{E}$ is positive definite and hence there exists a $p \times p$ lower triangular matrix $\mathbf{C}$ (the Cholesky decomposition) with non zero elements in the diagonal such that $\mathbf{E} = \mathbf{C}\mathbf{C}^t$. Furthermore, we have that:

1. $|\mathbf{C}| = |\mathbf{E}|^{\frac{1}{2}}$.
2. $(\mathbf{A}_1 + \mathbf{A}_2)^{-1} = \mathbf{A}_1^{-1}\mathbf{E}^{-1}\mathbf{A}_2^{-1}$.
3. $|\mathbf{A}_1 + \mathbf{A}_2| = |\mathbf{A}_1||\mathbf{A}_2||\mathbf{E}|$.

Using the identity,

$$\begin{aligned}&(\mathbf{y} - \mathbf{X}_1\theta_1)^t\mathbf{A}_1^{-1}(\mathbf{y} - \mathbf{X}_1\theta_1) + (\mathbf{y} - \mathbf{X}_2\theta_2)^t\mathbf{A}_2^{-1}(\mathbf{y} - \mathbf{X}_2\theta_2)\\ =&\mathbf{y}^t\mathbf{E}\mathbf{y} - 2\mathbf{y}^t(\mathbf{A}_1^{-1}\mathbf{X}_1\theta_1 + \mathbf{A}_2^{-1}\mathbf{X}_2\theta_2) + \theta_1^t\mathbf{X}_1^t\mathbf{A}_1^{-1}\mathbf{X}_1\theta_1 + \theta_2^t\mathbf{X}_2^t\mathbf{A}_2^{-1}\mathbf{X}_2\theta_2\\ =&\left[\mathbf{C}^t\mathbf{y} - \mathbf{C}^{-1}\mathbf{D}\right]^t\left[\mathbf{C}^t\mathbf{y} - \mathbf{C}^{-1}\mathbf{D}\right] + (\mathbf{X}_2\theta_2 - \mathbf{X}_1\theta_1)^t(\mathbf{A}_1 + \mathbf{A}_2)^{-1}(\mathbf{X}_2\theta_2 - \mathbf{X}_1\theta_1),\end{aligned}$$

along with the change of variables $\mathbf{u} = \mathbf{C}^t\mathbf{y} - \mathbf{C}^{-1}\mathbf{D}$, we conclude that

$$\begin{aligned}&\int_{\mathbb{R}^p}\mathbf{y}^t\mathbf{K}\mathbf{y}\prod_{i=1}^{2}N_p(\mathbf{y}|\mathbf{X}_i\theta_i, \mathbf{A}_i)d\mathbf{y} = \frac{\exp\left(-\frac{1}{2}(\mathbf{X}_2\theta_2 - \mathbf{X}_1\theta_1)^t(\mathbf{A}_1 + \mathbf{A}_2)^{-1}(\mathbf{X}_2\theta_2 - \mathbf{X}_1\theta_1)\right)}{(2\pi)^{\frac{p}{2}}|\mathbf{A}_1|^{\frac{1}{2}}|\mathbf{A}_2|^{\frac{1}{2}}|\mathbf{E}|^{\frac{1}{2}}}\\ &\times\int_{\mathbb{R}^p}\left[\mathbf{u}^t\mathbf{C}^{-1}\mathbf{K}(\mathbf{C}^t)^{-1}\mathbf{u} + \mathbf{D}^t(\mathbf{C}\mathbf{C}^t)^{-1}\mathbf{K}(\mathbf{C}\mathbf{C}^t)^{-1}\mathbf{D} + 2\mathbf{D}^t(\mathbf{C}\mathbf{C}^t)^{-1}\mathbf{K}(\mathbf{C}^t)^{-1}\mathbf{u}\right] N_p(\mathbf{u}|\mathbf{0}, \mathbf{I}_p)d\mathbf{u}\\ &= \left[\mathrm{tr}(\mathbf{C}^{-1}\mathbf{K}(\mathbf{C}^t)^{-1}) + \mathbf{D}^t(\mathbf{C}\mathbf{C}^t)^{-1}\mathbf{K}(\mathbf{C}\mathbf{C}^t)^{-1}\mathbf{D}\right] N_p(\mathbf{X}_2\theta_2|\mathbf{X}_1\theta_1, \mathbf{A}_1 + \mathbf{A}_2).\end{aligned}$$

To obtain the last equality we have used the fact that if $\mathbf{x}$ has a k dimensional distribution with mean vector μ and covariance matrix $\mathbf{\Sigma}$, then $E\left[\mathbf{x}^t\mathbf{A}\mathbf{x}\right] = tr\left[\mathbf{A}\mathbf{\Sigma}\right] + \mu^t\mathbf{A}\mu$. This proves the assertion. ■

The following corollary is an immediate consequence of lemma 2.

Corollary 2 *Let $\mathbf{K}$ be a $p \times p$ symmetric matrix and $\mathbf{X}$ a $p \times q$ matrix of rank q, $(q \leq p)$ such that $\mathbf{KX} = \mathbf{0}$. Then, if θ_i, $i = 1, 2$ are q dimensional vectors, we have*

$$\int_{\mathbb{R}^p} \mathbf{y}^t \mathbf{K} \mathbf{y} \prod_{i=1}^{2} N_p(\mathbf{y}|\mathbf{X}\theta_i, \sigma_i^2 \mathbf{I}_p) d\mathbf{y} = \frac{\sigma_2^2 tr(\mathbf{K}) |\mathbf{X}^t\mathbf{X}|^{-\frac{1}{2}}}{(2\pi\sigma_1^2)^{\frac{p-q}{2}} \left(1 + \frac{\sigma_2^2}{\sigma_1^2}\right)^{\frac{p-q+2}{2}}} N_q(\theta_2|\theta_1, (\sigma_1^2 + \sigma_2^2)(\mathbf{X}^t\mathbf{X})^{-1}).\Box$$

Theorem 1 *The intrinsic prior for α, σ_1 conditional on an arbitrary but fixed point γ_1, σ_0, turns out to be*

$$\pi^I(\alpha, \sigma_1|\gamma_1, \sigma_0) = N_{k+1}(\alpha|\gamma, (\sigma_0^2 + \sigma_1^2)\mathbf{W}^{-1}) \frac{1}{\sigma_0} \left(1 + \frac{\sigma_1^2}{\sigma_0^2}\right)^{-3/2}, \quad (3)$$

where $\gamma = (\mathbf{0}^t, \gamma_1^t)^t$ and $\mathbf{W} = \mathbf{Z}^t\mathbf{Z}$.

PROOF. Let $\mathbf{z}$ be a theoretical training sample of minimal size, and $\mathbf{Z}$ the corresponding matrix of regressors. Consider for $\mathbf{Z}$ the same partition as in the design matrix $\mathbf{X}$, that is $\mathbf{Z} = [\mathbf{Z}_0|\mathbf{Z}_1]$. Note that $\mathbf{Z}\gamma = \mathbf{Z}_1\gamma_1$, so that we can write

$$\pi^I(\alpha, \sigma_1|\gamma_1, \sigma_0) = \pi_1^N(\alpha, \sigma_1) E_{\mathbf{z}|\alpha,\sigma_1} \frac{N_{k+1}(\mathbf{z}|\mathbf{Z}\gamma, \sigma_0^2\mathbf{I}_{k+1})}{\int_{\mathbb{R}^k}\int_{\mathbb{R}^+} N_{k+1}(\mathbf{z}|\mathbf{Z}\alpha, \sigma_1^2\mathbf{I}_{k+1})\pi_1^N(\alpha, \sigma_1) d\alpha d\sigma_1}$$

$$= \frac{c_1}{\sigma_1^2} E_{\mathbf{z}|\alpha,\sigma_1} \frac{N_{k+1}(\mathbf{z}|\mathbf{Z}\gamma, \sigma_0^2\mathbf{I}_{k+1})}{\int_{\mathbb{R}^k}\int_{\mathbb{R}^+} N_{k+1}(\mathbf{z}|\mathbf{Z}\alpha, \sigma_1^2\mathbf{I}_{k+1}) \frac{c_1}{\sigma_1^2} d\alpha d\sigma_1}.$$

¿From corollary 2, we obtain

$$\begin{aligned}\pi^I(\alpha, \sigma_1|\gamma_1, \sigma_0) &= \frac{(2\pi)^{\frac{1}{2}}|\mathbf{W}|^{\frac{1}{2}}}{\sigma_1^2} E_{\mathbf{z}|\alpha,\sigma_1} \mathbf{z}^t\mathbf{Q}\mathbf{z} N_{k+1}(\mathbf{z}|\mathbf{Z}\gamma, \sigma_0^2\mathbf{I}_{k+1}) \\ &= \frac{(2\pi)^{\frac{1}{2}}|\mathbf{W}|^{\frac{1}{2}}}{\sigma_1^2} \int_{\mathbb{R}^{k+1}} \mathbf{z}^t\mathbf{Q}\mathbf{z} N_{k+1}(\mathbf{z}|\mathbf{Z}\gamma, \sigma_0^2\mathbf{I}_{k+1}) N_{k+1}(\mathbf{z}|\mathbf{Z}\alpha, \sigma_1^2\mathbf{I}_{k+1}) d\mathbf{z},\end{aligned}$$

where $\mathbf{Q} = \mathbf{I}_{k+1} - \mathbf{Z}\mathbf{W}^{-1}\mathbf{Z}^t$. The result follows from corollary 2 and the fact that tr$(\mathbf{Q}) = 1$. ■

For the particular case of testing the null hypothesis $H_0 : \alpha = \mathbf{0}$ versus $H_1 : \alpha \neq \mathbf{0}$, the intrinsic prior for parameters α and σ_1, conditional on σ_0 turns out to be

$$\pi^I(\alpha, \sigma_1|\sigma_0) = N_k(\alpha|\mathbf{0}, (\sigma_0^2 + \sigma_1^2)\mathbf{W}^{-1}) \frac{1}{\sigma_0} \left(1 + \frac{\sigma_1^2}{\sigma_0^2}\right)^{-3/2}.$$

It is interesting to note that in the intrinsic prior above, the marginal density of the regression coefficient vector α is centered at the null, a claim made by Morris [12] for sharp null hypothesis.

Remark 1 From equation 3, it follows that the marginal intrinsic prior for σ_1 conditional on γ_1, σ_0 is the square root of an Inverted-Beta-2 density with parameters 1/2, 1 and σ_0^2 (Raiffa and Schlaifer, [13], p. 221). The marginal distribution of α conditional on γ_1, and σ_0 is an elliptical multivariate distribution with mean vector γ. However, second and higher order moments do not exist since the mean of the mixing distribution is infinite. This implies that the marginal conditional intrinsic prior for α has very heavy tails as expected for a default prior. ■

3. Design considerations

Matrix $\mathbf{W}^{-1}$ in 3 depends on the theoretical regressors of a training sample of size $k+1$. A way to asses $\mathbf{W}^{-1}$ is by using the underlying idea in the origin of the Arithmetic Intrinsic Bayes Factor (Berger and Pericchi [1]) of averaging over all possible training samples of minimal size contained in the observed sample. This would give the matrix

$$\mathbf{W}^{-1} = \frac{1}{L}\sum_{\ell=1}^{L}(\mathbf{Z}^t(\ell)\mathbf{Z}(\ell))^{-1},$$

where $\{\mathbf{Z}(\ell), \ell = 1, \ldots, L\}$ is the set of all submatrices of $\mathbf{X}$ of order $(k+1)\times k$ such that $|\mathbf{Z}^t(\ell)\mathbf{Z}(\ell)| \neq \mathbf{0}$.

4. Bayes factor for intrinsic priors

For the data $\mathbf{D}$, the Bayes factor for testing H_0 *versus* H_1 with the intrinsic priors $\pi_0^N(\gamma_1,\sigma_0), \pi^I(\alpha,\sigma_1)$, say

$$B_{10}(\mathbf{D}) = \frac{\int N_n(\mathbf{y}|\mathbf{X}\alpha, \sigma_1^2\mathbf{I}_n)\pi_1^I(\alpha,\sigma_1)d\alpha d\sigma_1}{\int N_n(\mathbf{y}|\mathbf{X}_1\gamma_1, \sigma_0^2\mathbf{I}_n)\pi_0^N(\gamma_1,\sigma_0)d\gamma_1 d\sigma_0},$$

is given in theorem 2. First, we consider a previous result.

Lemma 3 *Let $\mathbf{X}$ be a $p\times q$ matrix of rank q ($q \le p$). If θ_1 and θ_2 are vectors of length q and $\mathbf{A}_i$, $i = 1,2$ are positive definite matrices, then*

$$\int_{\mathbb{R}^q} N_p(\mathbf{y}|\mathbf{X}\theta_1, \mathbf{A}_1)N_q(\theta_1|\theta_2, \mathbf{A}_2)d\theta_1 = N_p(\mathbf{y}|\mathbf{X}\theta_2, \mathbf{A}_1 + \mathbf{X}\mathbf{A}_2\mathbf{X}^t).$$

PROOF. Consider the identity

$$\begin{aligned}
&(\mathbf{y}-\mathbf{X}\theta_1)^t\mathbf{A}_1^{-1}(\mathbf{y}-\mathbf{X}\theta_1) + (\theta_1-\theta_2)^t\mathbf{A}_2^{-1}(\theta_1-\theta_2)\\
=&(\theta_1-\theta_2)^t\mathbf{B}(\theta_1-\theta_2) + (\mathbf{y}-\mathbf{X}\theta_2)^t\mathbf{A}_1^{-1}(\mathbf{y}-\mathbf{X}\theta_2) - 2(\mathbf{y}-\mathbf{X}\theta_2)^t\mathbf{A}_1^{-1}\mathbf{X}(\theta_1-\theta_2),
\end{aligned}$$

where $\mathbf{B} = \mathbf{A}_2^{-1} + \mathbf{X}'\mathbf{A}_1^{-1}\mathbf{X}$. Using this identity, we have

$$\begin{aligned}
&\int_{\mathbb{R}^q} N_p(\mathbf{y}|\mathbf{X}\theta_1, \mathbf{A}_1)N_q(\theta_1|\theta_2, \mathbf{A}_2)d\theta_1\\
&\quad= \frac{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\theta_2)^t\mathbf{A}_1^{-1}(\mathbf{y}+\mathbf{X}\theta_2)\right)}{(2\pi)^{\frac{p}{2}}|\mathbf{A}_1|^{\frac{1}{2}}|\mathbf{A}_2|^{\frac{1}{2}}|\mathbf{B}|^{\frac{1}{2}}}\int_{\mathbb{R}^q}\exp\left((\mathbf{y}-\mathbf{X}\theta_2)^t\mathbf{A}_1^{-1}\mathbf{X}\theta_1\right)N_q(\theta_1|\theta_2,\mathbf{B}^{-1})d\theta_1\\
&\quad= \frac{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\theta_2)^t\mathbf{A}_1^{-1}(\mathbf{y}+\mathbf{X}\theta_2)\right)}{(2\pi)^{\frac{p}{2}}|\mathbf{A}_1|^{\frac{1}{2}}|\mathbf{A}_2|^{\frac{1}{2}}|\mathbf{B}|^{\frac{1}{2}}}M_{\theta_1,\theta_2,\mathbf{B}^{-1}}(\mathbf{X}^t\mathbf{A}_1^{-1}(\mathbf{y}-\mathbf{X}\theta_2))\\
&\quad= \frac{\exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{X}\theta_2)^t\left[\mathbf{A}_1^{-1}-\mathbf{A}_1^{-1}\mathbf{X}\mathbf{B}^{-1}\mathbf{X}^t\mathbf{A}_1^{-1}\right](\mathbf{y}-\mathbf{X}\theta_2)\right)}{(2\pi)^{\frac{p}{2}}|\mathbf{A}_1|^{\frac{1}{2}}|\mathbf{A}_2|^{\frac{1}{2}}|\mathbf{B}|^{\frac{1}{2}}}\\
&\quad= N_p(\mathbf{y}|\mathbf{X}\theta_2, \mathbf{A}_1+\mathbf{X}\mathbf{A}_2\mathbf{X}^t).
\end{aligned}$$

The last equality follows from the standard matrix results:

1. If $\mathbf{P} = \mathbf{A} + \mathbf{CBD}$, then $\mathbf{P}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}\left(\mathbf{B}^{-1}+\mathbf{D}\mathbf{A}^{-1}\mathbf{C}\right)^{-1}\mathbf{D}\mathbf{A}^{-1}$.
2. Let $\mathbf{A}$ and $\mathbf{B}$ be matrices of dimension $p\times q$ and $q\times p$, respectively. Then, it follows that $|\mathbf{I}_p+\mathbf{AB}| = |\mathbf{I}_q+\mathbf{BA}|$. ■

Theorem 2 *The Bayes factor for comparing models*

$$M_0 : \{N_n(y|\mathbf{X}_1\gamma_1, \sigma_1^2\mathbf{I}_n), \pi_0^N(\gamma_1, \sigma_0)\}$$

and

$$M_1 : \{N_n(y|\mathbf{X}\alpha, \sigma_1^2\mathbf{I}_n), \pi_1^I(\alpha, \sigma_1)\},$$

turns out to be

$$B_{10}(\boldsymbol{D}) = |\mathbf{X}_1^t\mathbf{X}_1|^{1/2}(\mathbf{y}^t(\mathbf{I}_n - \mathbf{H}_1)\mathbf{y})^{(n-k_1+1)/2}I_0,$$

where $\mathbf{H}_1 = \mathbf{X}_1(\mathbf{X}_1^t\mathbf{X}_1)^{-1}\mathbf{X}_1^t$,

$$I_0 = \int_0^{\pi/2} \frac{d\varphi}{|\mathbf{A}(\varphi)|^{1/2}\,|\mathbf{B}(\varphi)|^{1/2}\,E(\varphi)^{(n-k_1+1)/2}},$$

and

$$\mathbf{B}(\varphi) = (\sin^2\varphi)\mathbf{I}_n + \mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t; \quad \mathbf{A}(\varphi) = \mathbf{X}_1^t\mathbf{B}(\varphi)^{-1}\mathbf{X}_1,$$
$$E(\varphi) = \mathbf{y}^t(\mathbf{B}(\varphi)^{-1} - \mathbf{B}(\varphi)^{-1}\mathbf{X}_1\mathbf{A}(\varphi)^{-1}\mathbf{X}_1^t\mathbf{B}(\varphi)^{-1})\mathbf{y}.$$

PROOF. From corollary 1, the marginal of the data under model M_0 is

$$m_0(\mathbf{D}) = \int_{\mathbb{R}^{k_1}}\int_{\mathbb{R}^+} N_n(\mathbf{y}|\mathbf{X}_1\gamma_1, \sigma_0^2\mathbf{I}_n)\pi_0^N(\gamma_1, \sigma_0)d\gamma_1 d\sigma_0$$
$$= \frac{c_0\Gamma\left(\frac{n-k_1+1}{2}\right)}{\sqrt{2}\pi^{\frac{n-k_1}{2}}|\mathbf{X}_1^t\mathbf{X}_1|^{\frac{1}{2}}\left(\mathbf{y}^t(\mathbf{I}_n - \mathbf{H}_1)\mathbf{y}\right)^{\frac{n-k_1+1}{2}}}.$$

The marginal of **D** under model M_1 is formally written as

$$m_1(\mathbf{D}) = \int_{\mathbb{R}^{k_1}}\int_{\mathbb{R}^+}\int_{\mathbb{R}^k}\int_{\mathbb{R}^+} N_n(\mathbf{y}|\mathbf{X}\alpha, \sigma_1^2\mathbf{I}_n)\pi_1^I(\alpha, \sigma_1|\gamma_1, \sigma_0)\pi_0^N(\gamma_1, \sigma_0)d\alpha d\gamma_1 d\sigma_1 d\sigma_0.$$

To evaluate the multiple integral we note that $\mathbf{X}\gamma = \mathbf{X}_1\gamma_1$, and from lemmas 3 and 1, we obtain

$$\int_{\mathbb{R}^k}\int_{\mathbb{R}^{k_1}} N_n(\mathbf{y}|\mathbf{X}\alpha, \sigma_1^2\mathbf{I}_n)N_k(\alpha|\gamma, (\sigma_0^2 + \sigma_1^2)\mathbf{W}^{-1})d\alpha d\gamma_1$$
$$= \int_{\mathbb{R}^{k_1}} N_n(\mathbf{y}|X_1\gamma_1, \mathbf{\Sigma})\,d\gamma_1 = \frac{\exp\left(-\frac{1}{2}\mathbf{y}^t\mathbf{A}_{\mathbf{\Sigma}}\mathbf{y}\right)}{(2\pi)^{\frac{n-k_1}{2}}\,|\mathbf{\Sigma}|^{\frac{1}{2}}|\mathbf{X}_1^t\mathbf{\Sigma}^{-1}\mathbf{X}_1|^{\frac{1}{2}}},$$

where $\mathbf{\Sigma} = \sigma_0^2\mathbf{I}_n + (\sigma_0^2 + \sigma_1^2)\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^t$ and

$$\mathbf{A}_{\mathbf{\Sigma}} = \mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{X}_1(\mathbf{X}_1^t\mathbf{\Sigma}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1^t\mathbf{\Sigma}^{-1}.$$

Therefore,

$$m_1(\mathbf{D}) = c_0\int_{\mathbb{R}^+}\int_{\mathbb{R}^+} \frac{\exp\left(-\frac{1}{2}\mathbf{y}^t\mathbf{A}_{\mathbf{\Sigma}}\mathbf{y}\right)}{(2\pi)^{\frac{n-k_1}{2}}\,|\mathbf{\Sigma}|^{\frac{1}{2}}|\mathbf{X}_1^t\mathbf{\Sigma}^{-1}\mathbf{X}_1|^{\frac{1}{2}}\sigma_0^3\left(1 + \frac{\sigma_1^2}{\sigma_0^2}\right)^{\frac{3}{2}}}d\sigma_0 d\sigma_1.$$

Changing σ_0 and σ_1 to polar coordinates, after some tedious but simple computations, renders

$$m_1(\mathbf{D}) = \frac{c_0\Gamma\left(\frac{n-k_1+1}{2}\right)}{\sqrt{2}\pi^{\frac{n-k_1}{2}}}\int_0^{\frac{\pi}{2}} \frac{d\varphi}{|\mathbf{B}(\varphi)|^{\frac{1}{2}}|\mathbf{A}(\varphi)|^{\frac{1}{2}}E(\varphi)^{\frac{n-k_1+1}{2}}}.$$

The Bayes factor $B_{10}(\mathbf{D})$ is now obtained as the ratio $m_1(\mathbf{D})/m_0(\mathbf{D})$ and this proves theorem 2. ■

For the particular case of testing the null hypothesis $H_0 : \alpha = \mathbf{0}$ versus $H_1 : \alpha \neq \mathbf{0}$, the Bayes factor turns out to be

$$B_{10}(\mathbf{D}) = (\mathbf{y}^t\mathbf{y})^{(n+1)/2} \int_0^{\pi/2} \frac{d\varphi}{|\mathbf{B}(\varphi)|^{1/2} E_1(\varphi)^{(n+1)/2}},$$

where

$$E_1(\varphi) = \mathbf{y}^t \mathbf{B}(\varphi)^{-1}\mathbf{y}.$$

5. Conclusions

Theorem 2 provides an automatic simple tool for the solution of the problem of testing general hypothesis of the form given by equation 1. In particular, this methodology can be applied to investigating the influence of some subsets of regressors in normal linear models, as is the case of matched pairs data, where the effectiveness of a treatment is under test.

Simulation results show that the performance of the Bayes Factor for intrinsic priors is very satisfactory (Girón, *et al.*, [5]). Further, this methodology is easily interpretable, takes into account the sample size and the design matrix automatically, which represents an important improvement over existing competing methods.

Acknowledgement. This paper has been partially supported by the Ministerio de Ciencia y Tecnología under grant projects PB97-1403-C03 and BEC2001-1982.

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.

[2] Berger, J. O. and Pericchi, L. R. (1997a). On the justification of default and intrinsic Bayes factor. In *Modeling and Prediction*, (J. C. Lee *et al.* eds.), New York; Springer-Verlag, 276–293.

[3] Berger, J. O. and Pericchi, L. R. (1997b). Accurate and stable Bayesian model selection: the median intrinsic Bayes factor. *Sankyà* B, **60**, 1–18.

[4] Berger, J. O. and Pericchi, L. R. (1998). On Criticism and Comparison of Default Bayes Factors for Model Selection and Hypothesis Testing. In *Proceedings of the Workshop on Model Selection* (W. Racugno ed.), Pitagora Ed., Bologna, 1–50.

[5] Girón, F. J., Martínez, M. L., Moreno, E. and Torres, F. (2003). Bayesian analysis of matched pairs in the presence of covariates. In *Bayesian Statistics*, **7** (J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith, eds.), to appear.

[6] Moreno, E. (1997). Bayes Factor for Intrinsic and Fractional Priors in Nested Models: Bayesian Robustness, in L_1*-Statistical Procedures and Related Topics*, (D. Yadolah ed.), Hayward, CA: Institute of Mathematical Statistics, **29**, 257–270.

[7] Moreno, E., Bertolino, F. and Racugno, W. (1998). An Intrinsic Limiting Procedure for Model Selection and Hypotheses Testing. *J. Amer. Statist. Assoc.*, **93**, 1451–1460.

[8] Moreno, E., Bertolino, F. and Racugno, W. (1999). Default Bayesian Analysis of the Behrens-Fisher Problem. *J. Statist. Plann. Inference*, **81**, 323–333.

[9] Moreno, E, Bertolino, F. and Racugno, W. (2000). Bayesian Model Selection Approach to Analysis of Variance under Heteroscedasticity. *J. Roy. Statist. Soc. Ser. D. The Statistician*, **46**, 1–15.

[10] Moreno, E. and Liseo, B. (2002). Default priors for testing the number of components of a mixture. *J. Statist. Plann. Inference*, to appear.

[11] Moreno, E., Torres, F. and Casella, G. (2002). Testing equality of regression coefficients in heteroscedastic normal regression models, *Submitted*.

[12] Morris, C. (1987). Discussion to Casella and Berger. *J. Amer. Statist. Assoc.*, **82**, 106– 111.

[13] Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*, M.I.T. Press; Cambridge, Mass.

E. Moreno
Dpto. de Estadística e I.O.
Facultad de Ciencias
Universidad de Granada
Spain
emoreno@ugr.es

F. J. Girón
Dpto. de Estadística e I.O.
Facultad de Ciencias
Universidad de Málaga
Spain
fj_giron@uma.es

F. Torres
Dpto. de Estadística e I.O.
Facultad de Ciencias
Universidad de Granada
Spain
fdeasis@ugr.es