



Available online at www.sciencedirect.com



C. R. Acad. Sci. Paris, Ser. I 340 (2005) 911–914



<http://france.elsevier.com/direct/CRASSI/>

Statistics

Consistency of Bayes factors for intrinsic priors in normal linear models

Elías Moreno ^a, F. Javier Girón ^b

^a *Departamento de Estadística e I. O., Universidad de Granada, Fuentenueva s/n, 18071 Granada, Spain*

^b *Departamento de Estadística e I. O., Universidad de Málaga, Campus de Teatinos s/n, 29071 Málaga, Spain*

Received 7 December 2004; accepted 18 April 2005

Available online 14 June 2005

Presented by Paul Deheuvels

Abstract

The Jeffreys–Lindley paradox refers to the well-known fact that a sharp null hypothesis on the normal mean parameter is always accepted when the variance of the conjugate prior goes to infinity, thus implying that the resulting Bayesian procedure is not consistent, and that some limiting forms of proper prior distributions are not necessarily suitable for testing problems. Intrinsic priors, which are limits of proper priors, have been proved to be extremely useful for testing problems, and, in particular, for testing hypothesis on the regression coefficients of normal linear models. This Note shows the consistency of the Bayes factors when using intrinsic priors for normal linear models under very mild conditions on the design matrix. **To cite this article:** *E. Moreno, F.J. Girón, C. R. Acad. Sci. Paris, Ser. I 340 (2005).*

© 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Résumé

Consistance des facteurs Bayes pour des lois a priori intrinsèques dans des modèles linéaires Gaussiens. Le paradoxe de Jeffreys–Lindley fait référence au fait bien connu qu’une hypothèse nulle sur la paramètre de moyenne d’une loi Gaussienne, qui est concentrée autour d’une valeur donnée est toujours acceptée, lorsque la variance de la loi à priori conjuguée tend vers l’infini, ce qui implique que la procédure Bayésienne associée n’est pas consistante, et que les lois à priori limites de distributions de probabilités, ne sont pas nécessairement appropriées pour des problèmes de tests d’hypothèse. Les lois à priori intrinsèques, qui sont elles-même limites de distributions de probabilité, se sont révélées être très utiles pour des problèmes de tests d’hypothèse, et en particulier, pour les tests concernant les coefficients de régression de modèles linéaires Gaussiens. Ce Note prouve la consistance des facteurs Bayes lorsque des lois à priori intrinsèques sont utilisées, dans des modèles linéaires Gaussiens, avec des conditions très faibles sur la matrice d’expérience. **Pour citer cet article :** *E. Moreno, F.J. Girón, C. R. Acad. Sci. Paris, Ser. I 340 (2005).*

© 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.

E-mail addresses: emoreno@ugr.es (E. Moreno), fj_giron@uma.es (F.J. Girón).

1631-073X/\$ – see front matter © 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.
doi:10.1016/j.crma.2005.05.001

1. Introduction

Suppose that data $\mathbf{x} = (x_1, \dots, x_n)$ come from a density $f(x|\mu)$ where μ is unknown, and it is desired testing that μ has a specific value μ_0 . The Bayesian approach to this problem is to compare the models

$$M_0: f(x|\mu_0), \quad \text{and} \quad M_1: \{f(x|\mu), \pi(\mu)\},$$

where $\pi(\mu)$ is a probability density describing our prior belief on the unknown parameter μ . We also need to assign a prior probability to each model, i.e., $P(M_0) = p_0$ and $P(M_1) = 1 - p_0$, and a loss function to measure the relative importance of making wrong decisions. Typically, a loss of the form $0 - 1$ is considered. Then, the optimal Bayesian decision is that of choosing model M_0 when the posterior probability $P(M_0|\mathbf{x})$ is greater than $1/2$.

The posterior probability of M_0 is given by

$$\Pr(M_0|\mathbf{x}) = \frac{B_{01}(\mathbf{x})}{B_{01}(\mathbf{x}) + (1 - p_0)/p_0},$$

where $B_{01}(\mathbf{x}) = f(\mathbf{x}|\mu_0) / \int f(\mathbf{x}|\mu)\pi(\mu) d\mu$ is the Bayes factor of model M_0 versus M_1 .

A well-known result, called the Jeffreys–Lindley paradox (see Lindley [2]), states that for testing a point null or sharp hypothesis on the mean of a normal distribution the posterior probability of the null tends to one when the variance of the conjugate prior tends to infinity, regardless the data. Furthermore, when the prior variance goes to infinity with an order $O(\exp\{n^\alpha\})$, $\alpha \geq 1$, the Bayesian approach is not consistent. As Robert [6] has pointed out this is not a mathematical paradox since the prior sequence is giving less and less mass to any neighborhood of the null point as the prior variance goes to infinity. However, an important consequence of the paradox is that some limiting forms of proper priors might not be suitable for testing problems as they could provide inconsistency of the corresponding Bayes factor.

On the other hand, an improper prior $\pi(\mu)$ can be rescaled to $c\pi(\mu)$, where the constant c can be arbitrarily chosen, so that the Bayes factor is not uniquely defined. Bayes factors for *intrinsic* priors (Berger and Pericchi [1]) were introduced to avoid this disturbing fact. *Intrinsic* priors are defined as the limit of a proper *restricted intrinsic* priors sequence (Moreno et al. [3]).

However, the consistency of the Bayes factor for intrinsic priors has not been established although they are limit of proper priors, and the Jeffreys–Lindley paradox might occur. In this Note, we prove that in linear models, and under very mild conditions on the design matrix, the consistency of the Bayes factor can be asserted. For readability of this Note, in Section 2 we briefly describe the formulas of the intrinsic priors and the Bayes factor for linear models.

2. Intrinsic priors and Bayes factors

Suppose that Y represents an observable random variable depending on X_1, X_2, \dots, X_k a set of k explanatory covariates through the normal linear model

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + \varepsilon, \quad \varepsilon \sim N(\cdot|0, \sigma^2).$$

Given n observations of Y , say $\mathbf{y} = (y_1, \dots, y_n)^t$, the likelihood function of (α, σ) is $N_n(\mathbf{y}|\mathbf{X}\alpha, \sigma^2\mathbf{I}_n)$, where the design matrix \mathbf{X} is assumed to be of rank k ($k \leq n$). Consider the partition of α as $\alpha^t = (\alpha_0^t, \alpha_1^t)$ and the corresponding partition of the columns of $\mathbf{X} = (\mathbf{X}_0|\mathbf{X}_1)$, so that \mathbf{X}_0 is of dimensions $n \times k_0$ and \mathbf{X}_1 is $n \times k_1$, where $k_1 = k - k_0$. We want to test whether the set of potential explanatory variables can be reduced to a tentative subset (X_{k_0+1}, \dots, X_k) , that is, we want to test $H_0: \alpha_0 = \mathbf{0}$, versus $H_1: \alpha_0 \neq \mathbf{0}$. A default Bayesian setting for this testing problem is that of choosing between the Bayesian models

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

and

$$M_1: N_n(\mathbf{y}|\mathbf{X}\alpha, \sigma^2\mathbf{I}_n), \pi_1^D(\alpha, \sigma) = \frac{c}{\sigma},$$

where π_1^D denotes the corresponding reference prior distribution for the regression coefficients and the standard error. These priors are improper, and the constants c_0 and c are arbitrary so that they cannot be used for solving the testing problem. A way to circumvent this problem is through the use of intrinsic priors.

Application of the standard intrinsic prior methodology to this problem (see Moreno, Girón and Torres [4]) shows that the intrinsic prior distribution for the parameters (α, σ) , conditional on a fixed point of the null (γ_1, σ_0) , turns out to be

$$\pi^I(\alpha, \sigma|\gamma_1, \sigma_0) = \frac{2}{\pi\sigma_0(1 + \sigma^2/\sigma_0^2)} N_k(\alpha|\tilde{\gamma}_1, (\sigma_0^2 + \sigma^2)\mathbf{W}^{-1}),$$

where $\tilde{\gamma}_1' = (\mathbf{0}', \gamma_1')$, and $\mathbf{W}^{-1} = n/(k + 1) \cdot (\mathbf{X}'\mathbf{X})^{-1}$. This matrix resembles the covariance matrix of Zellner’s g -priors, although \mathbf{W}^{-1} does not contain any tuning parameter.

Some lengthy algebraic manipulations show that the Bayes factor for the intrinsic priors $B_{01}(\mathbf{y}, \mathbf{X})$ can be written as

$$B_{01}(\mathbf{y}, \mathbf{X}) = \left(\frac{2(k + 1)^{k_0/2}}{\pi} \int_0^{\pi/2} \frac{\sin^{k_0} \varphi (n + (k + 1) \sin^2 \varphi)^{(n-k)/2}}{(n\mathcal{B}_n + (k + 1) \sin^2 \varphi)^{(n-k_1)/2}} d\varphi \right)^{-1}, \tag{1}$$

where \mathcal{B}_n is the statistic

$$\mathcal{B}_n = \frac{\mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y}}{\mathbf{y}'(\mathbf{I} - \mathbf{H}_1)\mathbf{y}},$$

\mathbf{I} is the $n \times n$ identity matrix, $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, and $\mathbf{H}_1 = \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1$.

3. Consistency of the Bayes factor for intrinsic priors

The weak consistency of the posterior probability $\Pr(M_0|n, \mathcal{B}_n)$ under the null and the alternative is proven in the next theorem.

Theorem 3.1. Under H_0 , $\Pr(M_0|n, \mathcal{B}_n) \xrightarrow{P} 1$. Further, if the limit

$$\mathbf{S} = \lim_{n \rightarrow \infty} \frac{\mathbf{X}'_0(\mathbf{I} - \mathbf{H}_1)\mathbf{X}_0}{n}$$

is a finite positive definite matrix then, under H_1 , $\Pr(M_0|n, \mathcal{B}_n) \xrightarrow{P} 0$.

Proof. The integrand in Eq. (1) can be written as

$$\begin{aligned} & \exp \left[k_0 \log \sin \varphi + \frac{n - k}{2} \log \left(1 + \frac{(k + 1) \sin^2 \varphi}{n} \right) \right] \\ & \times \exp \left[\frac{n - k_1}{2} \log \left(1 + \frac{(k + 1) \sin^2 \varphi}{n\mathcal{B}_n} \right) - \frac{k_0}{2} \log n - \frac{n - k_1}{2} \log \mathcal{B}_n \right]. \end{aligned} \tag{2}$$

The proof of the first part proceeds as follows: the first factor in (2) converges to $\exp[k_0 \log \sin \varphi + (k + 1)/2 \sin^2 \varphi]$ as $n \rightarrow \infty$. Under H_0 , \mathcal{B}_n follows a beta distribution with parameters $(n - k)/2$ and $k_0/2$; therefore, $\mathcal{B}_n \xrightarrow{P} 1$, so that the first random term in the exponent of (2) converges in probability to $-(k + 1)/2 \sin^2 \varphi$.

On the other hand, $-(n - k_1)/2 \log \mathcal{B}_n$ converges in distribution to a random variable following a Gamma distribution with parameters $k_0/2$ and 1. Thus, the second factor in (2), and consequently the integrand, converges in probability to 0 as the leading term is $\exp(-k_0/2 \log n)$. Now, using result (xiii), pp. 124, in Rao [5], and the continuity of the Bayes factor we get that, in probability, the Bayes factor $B_{01}(n, \mathcal{B}_n)$ converges to ∞ and, consequently, the posterior probability $\Pr(M_0|n, \mathcal{B}_n)$ converges to 1, whatever the value of the prior model probability p_0 .

Now we prove consistency under the alternative. Under H_1 , the statistic $1 - \mathcal{B}_n$ follows a noncentral beta distribution with parameters $k_0/2$, $(n - k)/2$ and noncentrality parameter $\delta = \alpha_0^t \mathbf{X}_0^t (\mathbf{I}_n - \mathbf{H}_1) \mathbf{X}_0 \alpha_0 / \sigma^2$. Then, it can be proved that

$$\mathcal{B}_n \xrightarrow{P} \left(1 + \frac{1}{\sigma^2} \alpha_0^t \mathbf{S} \alpha_0 \right)^{-1},$$

a value which is strictly smaller than 1 because $\alpha_0^t \mathbf{S} \alpha_0 > 0$ as \mathbf{S} is positive definite. This implies that the integrand (2) converges in probability to infinity as now the leading term $-(n - k_1)/2 \log \mathcal{B}_n$ is of order $O(n)$. Now, using again the continuity of the Bayes factor, we finally get that $\Pr(M_0|n, \mathcal{B}_n)$ tends to 0. \square

Remark 1. Using Zellner's g -prior the resulting Bayes factor is consistent under the alternative but it is not under the null. However, if g is of order $O(n^\alpha)$ with $\alpha > 0$, then it is consistent under the null and the alternative.

Corollary 3.2. *The Jeffreys–Lindley paradox does not hold in the normal model when using intrinsic priors.*

Acknowledgements

This paper has been partially supported by the *Ministerio de Ciencia y Tecnología* under grant SEJ2004-02447.

References

- [1] J.O. Berger, L.R. Pericchi, The intrinsic Bayes factor for model selection and prediction, *J. Amer. Statist. Assoc.* 91 (1996) 109–122.
- [2] D.V. Lindley, A statistical paradox, *Biometrika* 44 (1957) 187–192.
- [3] E. Moreno, F. Bertolino, W. Racugno, An intrinsic limiting procedure for model selection and hypotheses testing, *J. Amer. Statist. Assoc.* 93 (1998) 1451–1460.
- [4] E. Moreno, F.J. Girón, F. Torres, Intrinsic priors for hypotheses testing in normal regression models, *Rev. Acad. Cienc. Ser. A Mat.* 97 (2003) 53–61.
- [5] C.R. Rao, *Linear Statistical Inference and its Applications*, Wiley, New York, 1973.
- [6] C.P. Robert, A note on the Jeffreys–Lindley paradox, *Statist. Sinica* 3 (1993) 601–608.