Use of approximate marginal likelihood for model selection

L. Ventura

2002.18

Dipartimento di Scienze Statistiche
Università degli Studi
Via C. Battisti 241-243
35121 Padova

Dicembre 2002

Use of approximate marginal likelihood for model selection

Laura Ventura

Dipartimento di Scienze Statistiche,
Università degli Studi di Padova
Via C. Battisti 241 - 35121 Padova
e-mail: ventura@stat.unipd.it

Abstract

Optimal invariant tests for model discrimination exist when the two models under hypotheses represent scale-regression families. These tests are based on the ratio of the marginal likelihoods of the two families, based on the maximal invariant statistic, in order to eliminate the unknown parameters from the likelihood function. However, even in cases where these functions can in principle be found, it may be difficult to make the calculations required, since the resulting formula is expressed in terms of a multidimensional integral. In this paper a simple approximation to optimal invariant tests based on the Laplace formula is discussed. The main regularity condition required is that the maximum likelihood estimates of the scale and regression parameters exist.

1 Introduction

The use of marginal likelihoods in transformation models for parametric inference has been widely discussed (see, e.g. Pace and Salvan, 1997, chap. 7). In order to eliminate a set of nuisance parameters from the likelihood function, the basic idea is to consider a marginal likelihood based on the maximal invariant statistic. The resulting formula can be expressed in terms of an integral with respect to an invariant measure on the transformation parameter space. An important feature of marginal likelihoods is that they play a basic role in the construction of optimal invariant tests for model discrimination. To be specific, when the two models under hypotheses represent scale-regression families, there exist a most powerful invariant (MPI) test, based on the ratio of the marginal likelihoods of the two families. However, even in cases where these functions can in principle be found, it may be difficult to make the calculations required.

The purpose of this paper is to suggest an approximation to MPI tests, based on the Laplace formula for integrals. Its construction is similar to the approximation obained in Barndorff-Nielsen and Jupp (1988), and generalises the results in Pace and Salvan (1997, chap. 7) to regression-scale models. To apply the proposed approximation one only needs to be able to maximize a likelihood function and to evaluate the observed information matrix at the maximum. The resulting approximation is generally accuarte to order $O(n^{-1})$ and the main regularity condition required is that the maximum likelihood estimates (MLE) of the scale and regression parameters exist uniquely with probability one.

There is a close connection between the approximations proposed here and the approximations to posterior moments and marginal densities in Bayesian inference given by Tierney and Kadane (1986), and with approximate predictive likelihoods computed in Davison (1986).

2 MPI tests

A regression-scale model has the form $y = X\beta + \sigma\varepsilon$, where X is a fixed $n \times p$ matrix, $\beta \in \mathbb{R}^p$ an unknown regression coefficient, $\sigma > 0$ a scale parameter, and ε an n-dimensional vector of errors such that $(\varepsilon_1, \ldots, \varepsilon_n)$ is a random sample from a known density $p_\kappa(\cdot)$ on $\mathbb{R}$, where κ is a shpae parameter. Let $g_\kappa(\cdot) = -\log p_\kappa(\cdot)$ and $x_i^\top$ be the ith row of X. The log-likelihood for (β, σ) is

$$\ell(\beta, \sigma) = -n \log \sigma - \sum_{i=1}^{n} g_\kappa \left(\frac{y_i - x_i^\top \beta}{\sigma} \right) .$$

Here it is assumed that the MLE $(\hat{\beta}, \hat{\sigma})$ exists uniquely and is finite. The statistic $a = (a_1, \ldots, a_n)$, where $a_i = (y_i - x_i^\top \hat{\beta})/\hat{\sigma}$, $i = 1, \ldots, n$ is termed the sample configuration and it constitutes a maximal invariant statistic.

The marginal likelihood (LM) based on a is given by

$$L_m = \int_{\mathbb{R}^p} \int_{\mathbb{R}^+} \frac{1}{\sigma} \exp\{\ell_0(\beta, \sigma)\} d\beta d\sigma \ , \qquad (1)$$

i.e. the LM is expressible in terms of an integral with respect to a right invariant measure on the transformation parameter space (see e.g. Pace and Salvan, 1997, p. xxx). The LM plays a basic role in the construction of MPI tests for model discrimination. Consider the problem of assigning the density $p(\cdot)$ of y to one of the two separate regression-scale families, and suppose the assignment is done from the viewpoint of hypotheses testing. To be specific, let us denote by $\mathcal{F}_0$ and $\mathcal{F}_1$, with $\mathcal{F}_0 \cup \mathcal{F}_1 = \emptyset$, two possible regression-scale models. It is of interest in testing the null hypothesis

$$H_0 : p(y) \in \mathcal{F}_0 = \left\{ p(y; \beta, \sigma) = \frac{1}{\sigma} p_0 \left(\frac{y - x^\top \beta}{\sigma} \right) \right\}$$

against the alternative

$$H_1 : p(y) \in \mathcal{F}_1 = \left\{ p(y; \beta, \sigma) = \frac{1}{\sigma} p_1 \left(\frac{y - x^\top \beta}{\sigma} \right) \right\} .$$

Since both $\mathcal{F}_0$ and $\mathcal{F}_1$ represent regression-scale families, there exists a MPI test, obtainable through the ratio of marginal likelihoods, that is

$$T_{UMPI} = \frac{\int_{\mathbb{R}^p} \int_{\mathbb{R}^+} \frac{1}{\sigma} \exp\{\ell_1(\beta, \sigma)\} d\beta d\sigma}{\int_{\mathbb{R}^p} \int_{\mathbb{R}^+} \frac{1}{\sigma} \exp\{\ell_1(\beta, \sigma)\} d\beta d\sigma} \ . \qquad (2)$$

The former MPI test rejects the null hypothesis for large values of T_{UMPI}.

The MPI test in a closed form is given only for few particular situations. In general, MPI tests are unusable, since they do not admit a closed form expression. In practice, MPI tests can be applied only when it is possible to compute in a closed form the LM (1). But an obvious difficulty with (1) is the calculation of the multidimensional integral, and numerical approximation methods are needed. The natural idea in this setting is to consider suitable approximations. In particular, using the general approximations to LM investigated recently, we extend their use to MPI tests. The aim of this paper is to provide general and simple approximations to MPI tests, which are applicable in situations where maximum likelihood estimation is regular.

3 Laplace approximation

The two functions at the numerator and at the denominator in (2) can be interpretd as two marginal likelihoods for a index parameter κ for composite regression-scale models, where κ can take only values 0 and 1. An approximation on Laplace's formula to the marginal likelihood function for composite transformation models has been obtained in Barndorff-Nielsen and Jupp (1988) and in Pace and Salvan (1997, chap.) using a more familiar form for statisticians.

Let $r(\cdot)$ and $h(\cdot)$ be real valued functions defined on a subset $D \subseteq \mathbb{R}^d$ and assume that $r(\cdot)$ has a unique maximum in the interior of D at $\hat{\omega}$, where the gradient is zero and the Hessian $r''(\hat{\omega})$ is positive definite. Assume also that $h(\hat{\omega}) \neq 0$. The multiparameter version of the Laplace approximation is

$$I = \int \exp\{nr(\omega)\}h(\omega)d\omega = \frac{(2\pi)^{d/2}\exp\{nr(\hat{\omega})\}h(\hat{\omega})}{\sqrt{n}|i(\hat{\omega})|^{1/2}}\{1+O(n^{-1})\} \ , \qquad (3)$$

where $\hat{\omega}$ is the solution of the equation $\partial r(\omega)/\partial\omega = 0$, and $i(\hat{\omega})$ is minus the $(d\times d)$ matrix of second derivatives of $r(\omega)$ evaluated thereat. Later, the quantity n will be omitted from the integrand function, and it will not appear at the denominator of the approximation. In practice, this term will be absorbed in $r(\omega)$, and in particular n will be a measure of the information in the random sample, e.g. its size. Approximation (3) can be applied to (1) provided that the MLE $(\hat{\beta}, \hat{\sigma})$ for (β, σ) is an interior point of the integration region. In this case, $d = p+1$, $\omega = (\beta, \sigma)$, $h(\omega) = 1/\sigma$ and $r(\omega) = \ell(\omega)$. The resulting approximation for the LM is

$$\begin{aligned} LM &= LM_{app}\{1+O(n^{-1})\} \\ &= \frac{(2\pi)^{(p+1)/2}}{\hat{\sigma}}\exp\{\ell(\hat{\beta},\hat{\sigma})\}|j(\hat{\beta},\hat{\sigma})|^{-(1/2)}\{1+O(n^{-1})\} \ , \qquad (4) \end{aligned}$$

where $j(\cdot)$ denotes the observed information matrix evaluated at the MLE. As anticipated, the approximation (4) is easy to compute since it only requires the MLE for (β, σ) and the observed information matrix. Moreover, this approximation is accurate to order $O(n^{-1})$ and can be easily extended to the situation in which the maximum of the integrand function $\ell(\beta,\sigma)$ is on the boundary of the region of integration. Finally, it is possible to show that the approximation (4) is $O(n^{-1})$ equivalent to the modified profile likelihood function.

The approximate marginal likelihood (4) is very useful for testing hypothesis in the context of shape and regression-scale models, when inference about the shape parameter is of interest.

For separate models, when the expressions for the LM under H_0 and H_1 are computed, the approximation to the MPI tests (2) follows immediately. The approximate MPI test is

$$T_{app} = \frac{LM_{app}(1)}{LM_{app}(0)} \ . \tag{5}$$

It is possible to write the approximate MPI test (5) in terms of the likelihood ratio (LR) for the problem of testing $\mathcal{F}_0$ versus $\mathcal{F}_1$, namely

$$LR = \frac{sup_{(\beta,\sigma)} L(\beta,\sigma)_1}{sup_{(\beta,\sigma)} L(\beta,\sigma)_0} = \frac{L(\hat{\beta}_1,\hat{\sigma}_1)}{L(\hat{\beta}_1,\hat{\sigma}_1)} \ .$$

Functions $L(\hat{\beta}_i,\hat{\sigma}_i)$ are the maximized likelihood functions under the hypothesis H_i, $i=1,2$, and they can also be interpreted as the profile likelihoods $L_p(i)$ for the parameter i. Then, the LR can be defined as a profile likelihood ratio, since it can be written as $LR = L_p(1)/L_p(0)$. The approximate MPI test can be written in terms of the LR as

$$T_{app} = LR\sqrt{\frac{\hat{\sigma}_0\hat{j}_0}{\hat{\sigma}_1\hat{j}_1}} \ .$$

It may be oserved that, while LR is a profile likelihood ratio, the approximate MPI test statistic can be interpreted as a modified profile likelihood ratio.

4 An application

In order to illustrate the usefulness of the approximate MPI test, let us consider the problem of discriminating between the normal and the Student t distribution with $\nu = 2$ degrees of freedom. Sen and Srivastava (1990, pag. 32) consider a data set consisting of 26 observations on house prices. Among the variables examined are the selling price in thousands of dollars (y), the number of bedrooms (x_1), the floor space in square feet (x_2), the total number of rooms (x_3) and the front footage of lot in feet (x_4). The model can be written as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \sigma\varepsilon_i, \quad i = 1,\ldots,26, \ .$$

Preliminary analysis suggest to take ε_i to be standard Student's t with few degrees of freedom, to allow for longer tails and for extreme values.

A Monte Carlo comparison has been made of the power of the approsimate test with the power of the exact test (which requires numerical integration) and

n	T_m	T_{app}	LR
10	0.20	0.19	0.17
20	0.41	0.40	0.33
60	0.76	0.75	0.70
100	0.87	0.86	0.84

Table 1: Power of tests for discrimination between H_0 and H_1; $\alpha = 0.05$; number of replications 5000.

of the profile test for the problem of discriminating between the normal and the Student t_2 distribution.

Table 1 indicates that the approximate test T_{app} is nearly as good as the exact test T_m for the problem considered. Moreover T_{app} performs better than the profile LR.

5 Final remarks

In this paper we have discussed the use of the Laplace approximation in the context of testing hypotheses. The approximate tests are easy to compute, since they only require the MLEs and the observed information, and are accurate to order $O(n^{-1})$

Critical values for the approximate test may be determined by simulation. It also should be noted that, in general for all the approximate MPI tests, the obtainment of the null distribution does not represent a problem since Laplace aprosimation produces invariant expressions under regression and scale changes. Consequently, for each value of n, the simulation can be performed under a single distribution.

In order to compare the power of the approximate MPI tests and the power uf the LR, other examples have been considered. It is important to observe that, in all the situations considered, the approximate MPI tests performs better than the LR but the difference can be very small. The results of these simulations suggest the existance of a particular relationship between the approximate MPI test and the LR and, in the light of these results, it would be interesting to investigate if there exist a particular connection between the two tests.

ACKNOWLEDGMENTS. This work was partially supported by grants from Ministero dell'Università e della Ricerca Scientifica e Tecnologica, Italy.

References

BARNDROFF-NIELSEN, O.E., JUPP, P.E. (1988): Differential geometry, profile likelihood, L-sufficiency and composite transformation models. *Ann. Statist.*, *16*, 1009–1043.

DAVISON, A.C. (1986): Approximate predictive likelihood. *Biometrika, 732*, 323–332.

LAWLESS, J.F. (1982): *Statistical models and methods for life-time data analysis.* Wiley, New York.

PACE, L., SALVAN, A. (1997): *Principles of Statistical Inference.* World Scientific, Singapore.

SEN, A., SRIVASTAVA, M. (1990): *Regression Analysis: Theory, Methods and Applications.* Springer-Verlag, New York.

TIERNEY, L., KADANE, J.B. (1986): Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc., 81*, 82–87.