Likelihood based discrimination between separate scale and regression models

L. Pace, A. Salvan, L. Ventura

2003.16

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

Settembre 2003

Likelihood based discrimination between separate scale and regression models

L. Pace

Department of Statistics, University of Udine, Italy

A. Salvan and L. Ventura[1]

Department of Statistics, University of Padova, Italy

Abstract

The aim of this paper is to compare through simulation the likelihood ratio (LR) test with the most powerful invariant (MPI) test, and approximations thereof, for discriminating between two separate scale and regression models. The LR test as well as the approximate (first order) MPI test based on the leading term of the Laplace expansion for integrals are easy to compute. They only require the maximum likelihood estimates for the regression and scale parameters and the two observed informations. Even the approximate (second order) MPI test is not computationally heavy. On the contrary, the exact MPI test is expressed in terms of multidimensional integrals whose numerical evaluation appears reliable only in the two-dimensional and in the three-dimensional case. Two conclusions emerge in this paper. First, for scale and location models, exact (when computable) and approximate MPI tests are equivalent to the LR test in all the situations considered and for every sample size. This contrasts somehow

[1] *Address for correspondence*: Laura Ventura, Department of Statistics, University of Padova, Via C. Battisti 241, 35121 Padova, Italy, e-mail: `ventura@stat.unipd.it`.

with the prescription usually implied in the literature. Second, when the dimension of the regression parameter is a considerable fraction of a small or moderate sample size, the second order approximation to the MPI test clearly improves on the LR test, unlike the first order approximation.

Key words: Laplace expansion, likelihood ratio test, marginal likelihood, model selection, modified profile likelihood, most powerful invariant test, profile likelihood, scale and regression model.

1 Introduction

Consider observations $y = (y_1, \ldots, y_n)^\top$ from a scale and regression model of the form

$$y = X\beta + \sigma\varepsilon , \tag{1}$$

where X is a fixed $n \times p$ matrix of rank p, $\beta \in \mathbb{R}^p$ is a vector of unknown regression coefficients, $\sigma > 0$ is a scale parameter, and ε is an n-dimensional vector of errors such that $(\varepsilon_1, \ldots, \varepsilon_n)$ is a random sample from a density $p_\kappa(\cdot)$ on $\mathbb{R}$. Here the parameter $\kappa \in \mathcal{K} \subseteq \mathbb{N}$ indexes the set of possible error distributions. We will always consider regression models with intercept, i.e. where the first column of X is the unit vector.

Models of the form (1) are characterized by a linear predictor and possibly non-normal errors. Applications may be found in many areas such as survival analysis or quality control. Numerous error distributions $p_\kappa(\cdot)$ have been proposed for models of the form (1). Popular choices for $p_\kappa(\cdot)$ include the normal, Student's t, extreme value, logistic, and Cauchy distributions. Distributions on the positive real line include the exponential, lognormal, and gamma distributions.

Given a sample y from (1), one important inference problem is to select the error distribution, i.e. $\kappa \in \mathcal{K}$ corresponding to the best-fitting family

$$p_Y(y; \beta, \sigma, \kappa) = \prod_{i=1}^{n} \frac{1}{\sigma} p_\kappa\left(\frac{y_i - x_i^\top \beta}{\sigma}\right) , \tag{2}$$

where $x_i^\top$ is the i-th row of X, $i = 1, \ldots, n$. Attention is devoted here to discrimination between pairs of error distributions, i.e. models with $\mathcal{K} = \{0, 1\}$, supposing that the assignment is done from the viewpoint of hypothesis testing. To be specific, we consider the problem of testing

$$H_0 : \kappa = 0$$

against the alternative

$$H_1 : \kappa = 1 .$$

Under H_0 the error distribution is $p_0(\cdot)$ and under H_1 the error distribution is $p_1(\cdot)$. Tests for H_0 against H_1 are frequently referred to as tests for separate models, which are a special case of goodness-of-fit tests.

Several test statistics for separate models have been proposed for assessing the adequacy of the assumed error distribution (see e.g. Lawless, 1982, chap. 9, for a survey). Seminal papers in the field are Cox (1961) and Atkinson (1970). Recent contributions, from various perspectives, include Dass and Berger (2003), Yang (2003), Trottini and Spezzaferri (2002), and Cubedo and Oller (2002). The importance of sample size in model selection has been outlined by several authors (see e.g. Zucchini, 2000).

In the framework of scale and regression models, the two main frequentist procedures for separate models are the likelihood ratio (LR) test and the most powerful invariant (MPI) test. LR tests are based on the ratio of the likelihoods of the two families where the unknown parameters are substituted by their maximum likelihood estimates. LR test statistics can be easily computed in all the situations of practical interest. MPI tests are given by the ratio of the two marginal likelihoods based on the maximal invariant statistics, whose density does not depend on the unknown regression and scale parameters (see e.g. Fraser, 1979, chap. 6, and Barndorff-Nielsen and Cox, 1994, sec. 2.8). Marginal likelihoods can in principle be found, but it may be difficult to perform analytically or numerically the calculations required, since the resulting formula is expressed in terms of multidimensional integrals. However, the leading term of the Laplace expansion for integrals gives a simple first order approximation for these marginal likelihoods, which can be used to approximate MPI tests to first order. At the cost of some additional calculations, first order approximate MPI tests may be improved including the $O(n^{-1})$ term in the Laplace expansion. This gives second order approximate MPI tests.

The aim of this paper is to compare LR tests with exact and first and second order approximate MPI tests for model discrimination. Simulation results indicate that, in almost all the situations of practical interest, when $p = 1$, LR and both first order and second order approximate MPI tests give equivalent results for all n. Indeed, even the exact MPI test, computed analytically or numerically, gives the same results as the LR test in a number of problems considered. Hence, use of the simpler LR procedure is suggested for discriminating between separate scale and location families. When $p > 1$, LR and first order approximate MPI tests continue to give equivalent results. However, if second order approximate MPI tests are considered, a gain in power is achieved for small and moderate sample sizes.

The paper is organized as follows. Next section is devoted to LR tests. Marginal likelihood and MPI tests are discussed in Section 3. In Section 4 the approximation to MPI tests based on the Laplace expansion for integrals is discussed and its relation with LR tests is emphasized. Section 5 presents several simulation studies which compare LR and exact and first and second order approximate MPI tests. Concluding remarks are given in the last section.

2 Likelihood ratio procedures

The loglikelihood for (β, σ, κ) based on model (1) can be written as

$$\ell(\beta, \sigma, \kappa) = -n \log \sigma - \sum_{i=1}^{n} g_\kappa \left(\frac{y_i - x_i^\intercal \beta}{\sigma} \right) , \tag{3}$$

where $g_\kappa(\cdot) = -\log p_\kappa(\cdot)$. Here it is assumed that the maximum likelihood estimate (MLE) $(\hat{\beta}_\kappa, \hat{\sigma}_\kappa)$ of (β, σ) for fixed $\kappa \in \{0, 1\}$ exists uniquely and is finite (for a key condition see Burridge, 1981). The statistic

$$a = (a_1, \ldots, a_n) \ , \ \text{ with } a_i = \frac{y_i - x_i^\intercal \hat{\beta}_\kappa}{\hat{\sigma}_\kappa} \ , \ i = 1, \ldots, n, \tag{4}$$

termed the sample configuration, constitutes a maximal invariant statistic with respect to linear scale and regression transformations.

The loglikelihood function (3) plays a basic role in the construction of tests for model discrimination. The parameter (β, σ) in (3) is unknown, but it can be substituted with its restricted MLE in the two loglikelihoods. This gives the profile loglikelihood $\ell_p(\kappa) = \ell(\hat{\beta}_\kappa, \hat{\sigma}_\kappa, \kappa)$, $\kappa = 0, 1$. The LR test is then based on the profile likelihood ratio statistic

$$T_{LR} = \frac{\exp(\ell_p(1))}{\exp(\ell_p(0))} = \exp\left(\ell(\hat{\beta}_1, \hat{\sigma}_1, 1) - \ell(\hat{\beta}_0, \hat{\sigma}_0, 0)\right) . \tag{5}$$

The LR test rejects H_0 for large values of T_{LR}.

In the framework of scale and regression models, LR tests have some interesting features. First, they are invariant under the group of transformations that defines model (1), i.e. linear regression and scale transformations. Moreover, they are relatively easy to compute, requiring only the MLE of (β, σ) under H_0 and H_1. The main difficulty is that the exact distributions of T_{LR} under H_0 and H_1 are usually intractable. However, these distributions can be investigated through simulation.

3 MPI procedures

The use of marginal likelihoods for inference about index parameters in composite transformation models has been widely discussed (see e.g. Barndorff-Nielsen and Cox, 1994, chap. 2.8, and Pace and Salvan, 1997, chap. 7). The maximal invariant statistic (4) has a distribution that depends on κ only and inference about κ may be based on the induced model. The resulting distribution is the same if $(\hat{\beta}_\kappa, \hat{\sigma}_\kappa)$ is substituted by any equivariant estimator.

The marginal likelihood for κ based on the maximal invariant statistic a is given by (see e.g. Barndorff-Nielsen and Cox, 1994, sec. 2.8, and Fraser, 1979, chap. 2 and 6)

$$L_m(\kappa) = \int_{\mathbb{R}^+} \int_{\mathbb{R}^p} \sigma^{-1} \exp(\ell(\beta, \sigma, \kappa))\, d\beta d\sigma , \tag{6}$$

where $\ell(\beta, \sigma, \kappa)$ is the loglikelihood (3). The marginal likelihood (6) can be computed in a closed form only in a few special models. Examples thereof are given in Lawless (1972, 1973, 1978) for applications to Cauchy, logistic, Weibull and extreme value distributions and in Kappenman (1975) for the Laplace distribution, in the case of a scale and location model, and in Fraser (1976) for normal linear models.

The marginal likelihood (6) plays a basic role in the construction of MPI tests for model discrimination. The MPI test rejects the null hypothesis for large values of

$$T_{MPI} = \frac{\int_{\mathbb{R}^+} \int_{\mathbb{R}^p} \sigma^{-1} \exp(\ell(\beta, \sigma, 1)) d\beta d\sigma}{\int_{\mathbb{R}^+} \int_{\mathbb{R}^p} \sigma^{-1} \exp(\ell(\beta, \sigma, 0)) d\beta d\sigma} . \tag{7}$$

A closed form expression for (7) is seldom available. For the calculation of multidimensional integrals in (7) approximations are needed. In addition to numerical evaluation of the integrals, asymptotic approximations based on the Laplace expansion for integrals are useful.

4 Laplace expansion and MPI tests

In this section we discuss approximation to MPI tests based on the Laplace expansion for integrals (cfr. Barndorff-Nielsen and Cox, 1989, chap. 3, and Pace and Salvan, 1997, chap. 9). The leading term of the resulting approximation is generally accurate to order $O(n^{-1})$ and the main regularity condition required is that the restricted MLE's of (β, σ) exist uniquely with probability one (see Burridge, 1981). There is a close connection between the approximations considered here and the approximations to posterior moments and marginal densities in Bayesian inference given in Tierney and Kadane (1986), and with approximate predictive likelihoods computed in Davison (1986).

Let $r(\cdot)$ and $h(\cdot)$ be real valued functions defined on an open subset $\mathcal{D} \subseteq \mathbb{R}^d$, $d \geq 1$, and assume that $r(\cdot)$ has a unique maximum in the interior of $\mathcal{D}$ at $\hat{\omega}$, where the gradient is zero and the Hessian $\partial^2 r(\omega)/\partial\omega\partial\omega^\top$ is negative definite. Assume also that $h(\hat{\omega}) \neq 0$. The multivariate version of the Laplace expansion for integrals is

$$\int_{\mathcal{D}} h(\omega) \exp\left(nr(\omega)\right) d\omega = \frac{(2\pi)^{d/2} h(\hat{\omega}) \exp\left(nr(\hat{\omega})\right)}{\sqrt{n}|j(\hat{\omega})|^{1/2}} \left\{1 + C_1 + O(n^{-2})\right\} , \tag{8}$$

where $\hat{\omega}$ is the solution of the equation $\partial r(\omega)/\partial\omega = 0$, $j(\hat{\omega})$ is minus the $(d \times d)$ matrix of second derivatives of $r(\omega)$ evaluated thereat, i.e. $j(\omega) = -\partial^2 r(\omega)/\partial\omega\partial\omega^\top$, and C_1 is a correction term of order $O(n^{-1})$. Hereafter, the index n will be omitted from the integrand function, and it will not appear at the denominator of the approximation. In practice, this index will be absorbed into $r(\omega)$.

Approximations based on (8) can be applied to (6) provided that the MLE $(\hat{\beta}_\kappa, \hat{\sigma}_\kappa)$ of $\omega = (\beta, \sigma)$ for fixed κ is an interior point of the integration region. In this case, $d = p + 1$, $h(\omega) = 1/\sigma$ and $r(\omega) = \ell(\omega, \kappa)$. The resulting first order approximation for the marginal likelihood is

$$L_m^\dagger(\kappa) = \frac{(2\pi)^{(p+1)/2}}{\hat{\sigma}_\kappa |j(\hat{\beta}_\kappa, \hat{\sigma}_\kappa)|^{-(1/2)}} \exp\left(\ell(\hat{\beta}_\kappa, \hat{\sigma}_\kappa, \kappa)\right) = D(\kappa) \exp\left(\ell_p(\kappa)\right) , \tag{9}$$

where $j(\cdot)$ denotes the observed information matrix evaluated at the MLE and

$$D(\kappa) = \frac{(2\pi)^{(p+1)/2}}{\hat{\sigma}_\kappa |j(\hat{\beta}_\kappa, \hat{\sigma}_\kappa)|^{-(1/2)}} .$$

Note that

$$L_m(\kappa) = L_m^\dagger(\kappa) \left\{1 + O(n^{-1})\right\} .$$

From (9) it can be noted that the profile loglikelihood $\ell_p(\kappa)$ represents the leading term, of order $O(n)$, of the marginal loglikelihood. Moreover, $D(\kappa)$ is a correction term of order $O(1)$. As anticipated, approximation (9) is easy to compute. It only requires the MLE for (β, σ) and the observed information matrix evaluated thereat.

A problem with the approximation based on the leading term of the Laplace expansion (8) is that it can fail when the dimension of the integral is comparable with the asymptotic index n. This aspect has been previously pointed out by Shun and McCullagh (1995). In view of this, when p is large relative to n, it could be more appropriate to include the $O(n^{-1})$ correction term C_1 in the approximation of the marginal likelihood. Using index notation and Einstein summation convention, the expression of C_1 for (8) applied to (6) is

$$\begin{aligned} C_1 = C_1(\kappa) &= \frac{1}{24}\Big\{ 3\hat{j}^{rs}\hat{j}^{tu}\hat{j}^{vw}\hat{\ell}_{rst}\hat{\ell}_{uvw} + 2\hat{j}^{ru}\hat{j}^{sv}\hat{j}^{tw}\hat{\ell}_{rst}\hat{\ell}_{uvw} \\ &\quad + 3\hat{j}^{rs}\hat{j}^{tu}\hat{\ell}_{rstu} - \frac{12}{\hat{\sigma}}\hat{j}^{rs}\hat{j}^{t\sigma}\hat{\ell}_{rst} + \frac{2}{\hat{\sigma}^2}\hat{j}^{\sigma\sigma}\Big\} , \end{aligned} \quad (10)$$

where r, s, t, u, v, w range over $1, \ldots, p$, j^{rs} are the components of the inverse matrix of $j(\beta, \sigma)$, $\ell_{rst}\partial^3\ell(\omega,\kappa)/\partial\omega_r\partial\omega_s\partial\omega_t$, $\ell_{rstu} = \partial^4\ell(\omega,\kappa)/\partial\omega_r\partial\omega_s\partial\omega_t\partial\omega_u$, and the symbol "$\wedge$" over a likelihood quantity indicates evaluation at the MLE $(\hat{\beta}_\kappa, \hat{\sigma}_\kappa)$. The resulting second order approximation for the marginal likelihood is

$$L_m^*(\kappa) = L_m^\dagger(\kappa)\left\{1 + C_1(\kappa)\right\} , \quad (11)$$

so that

$$L_m(\kappa) = L_m^*(\kappa)\left\{1 + O(n^{-2})\right\} .$$

The first order approximate MPI test for $H_0 : \kappa = 0$ against $H_1 : \kappa = 1$ is

$$T^{\dagger}_{MPI} = \frac{L^{\dagger}_m(1)}{L^{\dagger}_m(0)} \; . \tag{12}$$

Note that

$$T^{\dagger}_{MPI} = T_{LR}\sqrt{\frac{\hat{\sigma}_0 \hat{\jmath}_1}{\hat{\sigma}_1 \hat{\jmath}_0}} \; , \tag{13}$$

and $T_{MPI} = T^{\dagger}_{MPI}\left\{1 + O(n^{-1})\right\}$. It may be observed that, while T_{LR} is a likelihood ratio statics from a profile likelihood, $T^{\dagger}_{MPI}$ can be interpreted as a likelihood ratio statistics from a modified profile likelihood. For large values of n, T_{LR} is the leading term of both $T^{\dagger}_{MPI}$ and T_{MPI}, so all these tests are expected to give the same conclusion. However, their behaviour in small or moderate samples may be different; see Section 5.

When approximation (11) for the marginal likelihood is used, we get the second order approximate MPI test

$$T^{*}_{MPI} = \frac{L^{*}_m(1)}{L^{*}_m(0)} \; , \tag{14}$$

with $T_{MPI} = T^{*}_{MPI}\left\{1 + O(n^{-2})\right\}$. Like T_{LR}, also $T^{\dagger}_{MPI}$ and T^{*}_{MPI} are invariant under linear regression and scale transformations.

5 Simulation studies

Several simulation studies have been performed in order to compare LR, exact MPI and first and second order approximate MPI tests for model discrimination. Both scale and location models and scale and regression models have been considered. In this section four examples will be presented in detail. The first example concerns a scale and location model. The second example discusses a simulation study for a scale and regression model based on a real data set, hence with a fixed value of p. Finally, Examples 3 and 4 allow both p and n to vary. Several other models have been considered in simulation studies not reported here, allowing n and p to vary in the ranges $5, \ldots, 200$ and $1, \ldots, 6$, respectively. The results emerging from the whole of these studies are that:

- T_{LR} is equivalent to the first order approximate MPI test, $T^{\dagger}_{MPI}$, for all values of n and p, except when p is large relative to n: in the latter case the first order Laplace expansion fails;
- T_{LR} and $T^{\dagger}_{MPI}$ are equivalent to the exact MPI test, T_{MPI}, for scale and location models, i.e. when $p = 1$, and for every sample size n;
- the second order approximate MPI test, T^{*}_{MPI}, gives better results than T_{LR} when $p > 1$ and the sample size n is small or moderate;
- for values of $p > 1$ instabilities in the numerical computation of the exact MPI test have been encountered; for this reason the exact MPI test computed using the `adapt` procedure of the library `integrate` of the `R` environment, T^{num}_{MPI}, has not been reported in Tables 2–4.

In all the simulation studies the power of the tests has been estimated fixing α to 0.05 and the number of replications to 10000. Critical values were also determined by simulation with 10000 replications. Our conclusions are insensitive to α.

Example 1: Discrimination between the normal and the Cauchy distributions. In this example, the reference models are scale and location families of the form (1), with $p = 1$ and with normal or Cauchy distributed errors. The LR, the exact and the first order approximate MPI tests have been considered. The LR test was initially studied by Dumonceaux, Antle and Haas (1973). T_{MPI} is the MPI test of normality against Cauchy alternatives derived by Franck (1981). Its expression is complicated. The statistic T_{MPI}^{num} is the MPI test computed numerically, using the `adapt` procedure of the library `integrate` of the `R` environment.

The results in Table 1 indicate that the first order approximate test $T_{MPI}^{\dagger}$ and the likelihood ratio test T_{LR} are as good as the exact test T_{MPI}. It must be noted that T_{LR} is easier to compute than $T_{MPI}^{\dagger}$ and T_{MPI}.

(Table 1 here)

Example 2: Discrimination between the normal and the Student's t_2 distributions. Sen and Srivastava (1990, pag. 32) consider a data set consisting of $n = 26$ observations on house prices. Among the variables examined are the selling price in thousands of dollars (y), the number of bedrooms (x_1) and the floor space in square feet (x_2). The model can be written as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \sigma\varepsilon_i, \quad i = 1, \ldots, 26 \; . \tag{15}$$

Preliminary analysis suggests to take ε_i to be standard Student's t_ν with few degrees of freedom, to allow for longer tails and for extreme values.

Let us consider the problem of discriminating between the normal and the Student's t_ν distribution with $\nu = 2$ degrees of freedom. The Monte Carlo study has been performed for four different values of n: $n = 26$ corresponds to the original data set, $n = 10$ to the first 10 observations and $n = 52$ and $n = 104$ to the situations with 2 and 4 replications, respectively, of the values of x_1 and x_2 in (15).

Again, the results in Table 2 indicate that the first order approximate test $T^{\dagger}_{MPI}$ is as good as T_{LR}. The same results were found also for other discrimination problems between two scale and regression families based on real data sets, such as Weibull versus gamma, logistic versus Cauchy, lognormal versus gamma, Weibull versus lognormal.

(Table 2 here)

Example 3: Discrimination between the logistic and the extreme value distributions. Here, the LR, the first and second order approximate MPI tests have been considered. The LR test for discriminating between the logistic and the extreme value distribution was considered in Berkson (1957). The Monte Carlo study has been performed for four different values of n ($n = 10, 20, 50, 100$) and allowing p to vary in the range $1, \ldots, 5$. The model considered is

$$y_i = \beta_0 + \beta_1 x_i + \ldots + \beta_{p-1} x_i^{p-1} + \sigma \varepsilon_i \ , \tag{16}$$

where, for each n, $(x_1, \ldots, x_n)$ is a random sample from a standard normal distribution.

The results in Table 3 indicate that $T^{\dagger}_{MPI}$ and T_{LR} behave very similarly for the problem considered, except when p is large relative to n. On the other hand, the second order approximate MPI test T^{*}_{MPI} gives a clear improvement over T_{LR} for $p > 1$ and small or moderate values of n. However, when $p = 1$, i.e. for scale and location models, LR and the first and second order approximate MPI tests are all equivalent for every sample size n.

(Table 3 here)

Example 4: Discrimination between the normal and the extreme value distributions. Let us consider the problem of discriminating between the normal and the extreme value distributions using LR test and first and second order approximate MPI tests. The Monte Carlo study has been performed as in Example 3, with a model of the form (16). However, only the marginal likelihood for the extreme value distribution had to be approximated. With normal errors, the marginal likelihood has indeed a closed form (see Fraser, 1979, chap. 6, and Szkutnik, 1988).

The results in Table 4 give the same indications as those in Table 3. $T^{\dagger}_{MPI}$ and T_{LR} are equivalent, except for $p = 4$ and $n = 10, 20$. The second order approximate MPI test gives appreciable gains in power for $p > 1$ and small or moderate values of n. Finally, when $p = 1$, LR and the first and second order approximate MPI tests are all equivalent for all the values of n.

(Table 4 here)

6 Final remarks

In this paper we have discussed the use of the LR test, the MPI test and of its approximations based on the Laplace expansion for model discrimination. The LR and the first order approximate MPI tests are easy to compute, since they only require the maximum likelihood estimates for the regression and scale parameters and the observed informations. Moreover, when $p > 1$ and the sample size is small or moderate, first order approximate

MPI test can be modified by including a higher-order correction term. Even the second order approximate MPI test is not computationally heavy, while the numerical evaluation of the exact MPI test appears reliable only in the two parameter and possibly in the three parameter models.

The main result of this paper is that in all the situations considered, the first order approximate MPI tests are equivalent or worse than the LR tests. In the light of this, for the problem of discriminating between two separate scale and regression models, LR procedures can be used safely when p is small and the sample size is moderate to large. Indeed, they give the same results as the approximate optimal procedures, with the advantage of simplicity. Moreover, when $p = 1$, i.e. for scale and location models, T_{LR} appears equivalent even to T_{MPI}.

Another finding is that, when p is a considerable fraction of a small or moderate n, the second order approximate MPI test $T^{\star}_{MPI}$ clearly improves on the LR test, unlike $T^{\dagger}_{MPI}$.

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

References

Atkinson, A. C. (1970), A method of discriminating between models (with discussion), *J. R. Statist. Soc.* B, **32**, 323–353.

Barndorff-Nielsen, O.E., Cox, D.R. (1989), *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, London.

Barndorff-Nielsen, O.E., Cox, D.R. (1994), *Inference and Asymptotics*, Chapman and Hall, London.

Berkson, J. (1957), Smoking and lung cancer: Some observations on two recent reports, *J. Amer. Statist. Assoc.*, **53**, 28–38.

Burridge, J. (1981), A note on maximum likelihood estimation for regression models using grouped data, *J. Roy. Statist. Soc.* B, **43**, 41–45.

Cox, D. R. (1961), Tests of separate families of hypotheses, *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, **1**, 105–123.

Cubedo, M., Oller, J.M. (2002), Hypothesis testing: a model selection approach, *J. Statist. Plan. Inf.*, **108**, 3–21.

Dass, S.C., Berger, J.O. (2003), Unified conditional frequentist and Bayesian testing of composite hypotheses, *Scand. J. Statist.*, **30**, 193–210.

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

Dumonceaux, R., Antle, C.E., Haas, G. (1973), Likelihood ratio test for discrimination between two models with unknown location and scale parameters, *Technometrics*, **15**, 19–27.

Franck, W.E. (1981), The most powerful invariant test of normal versus Cauchy with applications to stable alternatives, *J. Amer. Statist. Assoc.*, **76**, 1002–1005.

Fraser, D.A.S. (1976), Necessary analysis and adaptive inference, *J. Amer. Statist. Assoc.*, **71**, 99–110.

Fraser, D.A.S. (1979), *Inference and Linear Models*, McGraw Hill, New York.

Kappenman, R.F. (1975), Conditional confidence intervals for double exponential distribution parameters, *Technometrics*, **17**, 233–235.

Lawless, J.F. (1972), Conditional confidence interval procedures for the location and scale parameters of the Cauchy and logistic distributions, *Biometrika*, **59**, 377–386.

Lawless, J.F. (1973), Conditional versus unconditional confidence intervals for the parameters of the Weibull distribution, *J. Amer. Statist. Assoc.*, **68**, 665–669.

Lawless, J.F. (1978), Confidence interval estimation for the Weibull and extreme value distributions, *Technometrics*, **20**, 355–364.

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.

Severini, T.A. (2000), *Likelihood Methods in Statistics*, Oxford University Press, Oxford.

Shun, Z., McCullagh, P. (1995), Laplace approximation of high dimensional integrals, *J. R. Statist. Soc.* B, **57**, 749–760.

Szkutnik, Z. (1988), Most powerful invariant tests for binormality, *Ann. Statist.*, **16**, 292–301.

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

Trottini, M., Spezzaferri, F. (2002), A generalized predictive criterion for model selection, *Can. J. Statist.*, **30**, 79–96.

Yang, Y. (2003), Regression with multiple candidate models: selecting or mixing?, *Statistica Sinica*, **13**, to appear.

Zucchini, W. (2000), An introduction to model selection, *J. Math. Psychology*, **44**, 41–61.

n	T_{MPI}	T_{MPI}^{num}	$T_{MPI}^{\dagger}$	T_{LR}
10	0.67	0.65	0.67	0.66
20	0.92	0.93	0.92	0.92
50	0.98	0.98	0.98	0.98
100	0.99	0.99	0.99	0.99

Table 1: Power of tests for discrimination between normal and Cauchy distributions, with $p = 1$.

n	$T^{\dagger}_{MPI}$	T_{LR}
10	0.32	0.33
26	0.55	0.55
52	0.87	0.87
104	0.99	0.99

Table 2: Power of tests for discrimination between the normal and the Student t_2 distributions, with $p = 3$.

p	n	T^*_{MPI}	$T^\dagger_{MPI}$	T_{LR}
1	10	0.27	0.27	0.27
	20	0.49	0.49	0.49
	50	0.87	0.87	0.87
	100	0.99	0.99	0.99
2	10	0.28	0.18	0.19
	20	0.46	0.42	0.41
	50	0.83	0.81	0.81
	100	0.99	0.99	0.99
3	10	0.27	0.17	0.17
	20	0.47	0.39	0.39
	50	0.82	0.79	0.79
	100	0.99	0.98	0.98
4	10	0.33	0.11	0.13
	20	0.45	0.37	0.36
	50	0.80	0.77	0.77
	100	0.98	0.97	0.97
5	10	0.32	0.08	0.12
	20	0.38	0.28	0.35
	50	0.80	0.75	0.75
	100	0.97	0.97	0.97

Table 3: Power of tests for discrimination between the logistic and the extreme value.

p	n	T^*_{MPI}	$T^\dagger_{MPI}$	T_{LR}
1	10	0.18	0.18	0.18
	20	0.34	0.34	0.34
	50	0.79	0.79	0.79
	100	0.99	0.99	0.99
2	10	0.18	0.16	0.16
	20	0.32	0.28	0.27
	50	0.78	0.77	0.77
	100	0.99	0.99	0.99
3	10	0.13	0.12	0.12
	20	0.26	0.22	0.22
	50	0.77	0.76	0.76
	100	0.98	0.98	0.98
4	10	0.13	0.10	0.11
	20	0.25	0.20	0.22
	50	0.74	0.72	0.72
	100	0.98	0.98	0.98

Table 4: Power of tests for discrimination between the normal and the extreme value distributions.