



Department of Statistical Sciences
University of Padua
Italy

UNIVERSITÀ
DEGLI STUDI
DI PADOVA
DIPARTIMENTO
DI SCIENZE
STATISTICHE

A Neyman-Scott phenomenon in model discrimination

Luigi Pace

Department of Statistics
University of Udine
Italy

Alessandra Salvan, Laura Ventura

Department of Statistical Sciences
University of Padua
Italy

Abstract: The aim of this paper is to show through simulation that a Neyman-Scott phenomenon may occur in discriminating among separate stratified models. We focus on models which are scale families in each stratum. We consider traditional model selection procedures, such as the Akaike and Takeuchi information criteria, together with procedures based on the marginal likelihood and its Laplace approximation. We perform two simulation studies. Results indicate that, when the sample size in each stratum is fixed and the number of strata increases, correct selection probabilities for traditional model selection criteria may approach zero. On the other hand, model selection based on exact or approximate marginal likelihoods, that exploit invariance, can behave far better.

Keywords: Akaike's information criterion, Invariance, Laplace expansion, Marginal likelihood, Nuisance parameter, Profile likelihood, Stratified model, Takeuchi's information criterion.

Contents

1	Introduction	1
2	Likelihood procedures	2
3	Simulation results	4
4	Discussion	7

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
<http://www.stat.unipd.it>

Corresponding author:
A.Salvan
tel: +39 049 827 4168
alessandra.salvan@unipd.it
<http://www.stat.unipd.it/~salvan>

A Neyman-Scott phenomenon in model discrimination

Luigi Pace

Department of Statistics
University of Udine
Italy

Alessandra Salvan, Laura Ventura

Department of Statistical Sciences
University of Padua
Italy

Abstract: The aim of this paper is to show through simulation that a Neyman-Scott phenomenon may occur in discriminating among separate stratified models. We focus on models which are scale families in each stratum. We consider traditional model selection procedures, such as the Akaike and Takeuchi information criteria, together with procedures based on the marginal likelihood and its Laplace approximation. We perform two simulation studies. Results indicate that, when the sample size in each stratum is fixed and the number of strata increases, correct selection probabilities for traditional model selection criteria may approach zero. On the other hand, model selection based on exact or approximate marginal likelihoods, that exploit invariance, can behave far better.

Keywords: Akaike's information criterion, Invariance, Laplace expansion, Marginal likelihood, Nuisance parameter, Profile likelihood, Stratified model, Takeuchi's information criterion.

1 Introduction

Consider models for independent stratified observations of the form

$$Y_{ij} \sim p(y_{ij}; \psi, \lambda_i), \quad (1)$$

with $i = 1, \dots, q$ and $j = 1, \dots, m$, where q is the number of strata and m is the size of each stratum. Here ψ is a parameter indexing a set of k competing models, with $\psi \in \Psi = \{1, \dots, k\}$, $k \geq 2$. The parameter $\lambda = (\lambda_1, \dots, \lambda_q) \in \Lambda$ is a nuisance parameter, with λ_i allowing for detailed description of the i -th stratum. Note that Λ , assumed to be a subset of a Euclidean space, does not depend on ψ .

The inference problem to be considered, on the basis of a sample $y = (y_{11}, \dots, y_{qm})$ of size $n = mq$, with y_{ij} obeying (1), is to estimate ψ , i.e. to select one model among those available. The simplest way is through maximum likelihood. This amounts to select ψ maximising the profile likelihood (see e.g. Severini, 2000, Section 4.6).

Penalisations of maximised likelihood depending only on the dimension of λ and the sample size, such as the Akaike and the Bayesian information criteria, AIC (Akaike, 1973) and BIC (Schwarz, 1978), are equivalent to maximum likelihood for model (1). On the other hand, selection based on the Takeuchi information criterion, TIC (Takeuchi, 1976), that uses a data and model dependent penalisation, may differ from maximum likelihood. When the competing models are group families under the same group of transformations, selection can be based also on the marginal likelihood (see e.g. Quesenberry, 1985). The latter is expressed in terms of an integral. If not available in a closed form, the marginal likelihood can be approximated using Laplace expansion (see e.g. Ducharme and Frichot, 2003).

Models of form (1) have parameter whose dimension increases with the number of strata. Likelihood inference about ψ is expected to perform poorly when q is large relative to m . In particular, there are many well-known examples with ψ ranging in an interval, where, when q diverges and m is fixed, the Neyman-Scott phenomenon occurs, i.e. the maximum likelihood estimator of ψ is inconsistent (Neyman and Scott, 1948).

The aim of this paper is to show through simulation that a form of Neyman-Scott phenomenon may occur also in discriminating among separate stratified models. We focus on stratified models which are separate scale families in each stratum. We perform two simulation studies. In the first we consider selection between stratified exponential and half-normal models. In the second we study selection between stratified lognormal and inverse Gaussian models. The former selection problem generalises to stratified models the Example in Section 7.6.1 of Burnham and Anderson (2002). The latter problem generalises to stratified models the discrimination problem considered by Strupczewski *et al.* (2006) in the context of flood frequency analysis.

Results indicate that, when the sample size in each stratum is fixed and the number of strata increases, correct selection probabilities for traditional model selection criteria may approach zero. On the other hand, we find that model selection based on exact and approximate marginal likelihoods, that exploit invariance, gives far better results. Hence, if correct selection probabilities are considered important, model selection should be based on exact, or at least accurate, elimination of nuisance parameters.

The likelihood based selection procedures used in the simulation studies are reviewed in Section 2. Section 3 presents the simulation results. The final section contains a short discussion.

2 Likelihood procedures

Consider data y_{ij} generated from (1), with $i = 1, \dots, q$ and $j = 1, \dots, m$, where q is the number of strata and m is the size of each stratum. We assume in particular that y_{ij} are realisations of independent random variables Y_{ij} with density $p(y_{ij}; \psi, \lambda_i) = \lambda_i p(\lambda_i y_{ij}; \psi, 1)$, $\psi \in \{0, 1\}$, $\lambda_i > 0$. Here λ_i represents a scale parameter for the i -th

stratum. The full likelihood is

$$L(\psi, \lambda) = \prod_{i=1}^q L(\psi, \lambda_i),$$

where $L(\psi, \lambda_i) = \prod_{j=1}^m p(y_{ij}; \psi, \lambda_i)$. The profile likelihood for ψ is

$$L_P(\psi) = \prod_{i=1}^q L(\psi, \hat{\lambda}_{i\psi}),$$

where $\hat{\lambda}_{i\psi}$ is the maximum likelihood estimate of λ_i for a given ψ . We denote by $\ell_P(\psi) = \log L_P(\psi)$ the profile loglikelihood.

Under model (1), model selection based on maximization of $L_P(\psi)$ is equivalent to selection based on minimisation of Akaike's information criterion

$$\text{AIC} = 2(-\ell_P(\psi) + q).$$

AIC corresponds to the penalised profile loglikelihood

$$\ell_{\text{AIC}}(\psi) = \ell_P(\psi) - q. \quad (2)$$

Moreover, under model (1), model selection based on the profile likelihood is also equivalent to selection based on the Bayesian information criterion

$$\text{BIC} = 2(-\ell_P(\psi) + q \log n)$$

and on the corresponding penalised profile loglikelihood.

On the other hand, under model (1), Takeuchi's information criterion (see Burnham and Anderson, 2002, formula (7.38)) provides a different selection procedure, based on minimisation of

$$\text{TIC} = 2 \left[-\ell_P(\psi) + \sum_{i=1}^q j_{\lambda_i \lambda_i}(\psi, \hat{\lambda}_{i\psi})^{-1} \hat{\nu}_{\lambda_i, \lambda_i}(\psi, \hat{\lambda}_{i\psi}) \right],$$

where $j_{\lambda_i \lambda_i}(\psi, \lambda_i)$ is the observed information for λ_i in the i -th stratum for a given ψ , and

$$\hat{\nu}_{\lambda_i, \lambda_i}(\psi, \lambda_i) = \sum_{j=1}^m \ell_{\lambda_i}^{(ij)}(\psi, \lambda_i) \ell_{\lambda_i}^{(ij)}(\psi, \lambda_i),$$

with $\ell_{\lambda_i}^{(ij)}(\psi, \lambda_i) = \partial \log p(y_{ij}; \psi, \lambda_i) / \partial \lambda_i$. TIC corresponds to the penalised profile loglikelihood

$$\ell_{\text{TIC}}(\psi) = \ell_P(\psi) - \sum_{i=1}^q j_{\lambda_i \lambda_i}(\psi, \hat{\lambda}_{i\psi})^{-1} \hat{\nu}_{\lambda_i, \lambda_i}(\psi, \hat{\lambda}_{i\psi}). \quad (3)$$

The marginal likelihood based on the maximal invariant under stratum-wise scale transformations is given by

$$L_I(\psi) = \prod_{i=1}^q \int_0^{+\infty} \frac{1}{\lambda_i} L(\psi, \lambda_i) d\lambda_i. \quad (4)$$

The Laplace approximation for $\ell_I(\psi) = \log L_I(\psi)$ is

$$\ell_L(\psi) = \frac{q}{2} \log(2\pi) + \ell_P(\psi) - \sum_{i=1}^q \log \hat{\lambda}_{i\psi} - \frac{1}{2} \sum_{i=1}^q \log j_{\lambda_i \lambda_i}(\psi, \hat{\lambda}_{i\psi}). \quad (5)$$

Under model (1), $\ell_L(\psi)$ coincides with the modified profile loglikelihood of Barndorff-Nielsen (1983) and is invariant under interest respecting reparameterisations, so that it has the same expression if the parameters λ_i are substituted by $\sigma_i = 1/\lambda_i$.

3 Simulation results

Example 1: Selection between stratified exponential and half-normal models.

Suppose that, under the first model, where $\psi = 0$, Y_{ij} has an exponential density with mean $1/\lambda_i$, i.e.

$$p(y_{ij}; 0, \lambda_i) = \lambda_i \exp(-\lambda_i y_{ij}),$$

while under the second model, where $\psi = 1$, Y_{ij} has a half-normal distribution with mean $1/\lambda_i$, i.e.

$$p(y_{ij}; 1, \lambda_i) = \frac{2}{\pi} \lambda_i \exp(-y_{ij}^2 \lambda_i^2 / \pi).$$

Let $\bar{y}_i = (1/m) \sum_{j=1}^m y_{ij}$ and $\tilde{y}_i = (1/m) \sum_{j=1}^m y_{ij}^2$, $i = 1, \dots, q$. We have $\hat{\lambda}_{i\psi} = 1/\bar{y}_i$ if $\psi = 0$ and $\hat{\lambda}_{i\psi} = \sqrt{\pi/(2\tilde{y}_i)}$ if $\psi = 1$, so that the profile loglikelihood is

$$\ell_P(\psi) = \begin{cases} -m \sum_{i=1}^q \log(\bar{y}_i) - mq & \text{if } \psi = 0 \\ \frac{mq}{2} \log\left(\frac{2}{\pi}\right) - \frac{m}{2} \sum_{i=1}^q \log(\tilde{y}_i) - \frac{mq}{2} & \text{if } \psi = 1, \end{cases}$$

which is equivalent to $\ell_{AIC}(\psi)$ given by (2).

The marginal loglikelihood derived from (4) is

$$\ell_I(\psi) = \begin{cases} q \log \Gamma(m) - m \sum_{i=1}^q \log(m\bar{y}_i) & \text{if } \psi = 0 \\ q(m-1) \log 2 - \frac{qm}{2} \log \pi + q \log \Gamma\left(\frac{m}{2}\right) - \frac{m}{2} \sum_{i=1}^q \log(m\tilde{y}_i) & \text{if } \psi = 1. \end{cases}$$

For each model we can determine the likelihood quantities needed to compute $\ell_{TIC}(\psi)$ according to (3), and the Laplace approximation $\ell_L(\psi)$ of the marginal loglikelihood $\ell_I(\psi)$ according to (5). We have

$$j_{\lambda_i \lambda_i}(\psi, \hat{\lambda}_{i\psi}) = \begin{cases} m\bar{y}_i^2 & \text{if } \psi = 0 \\ 4m\tilde{y}_i/\pi & \text{if } \psi = 1, \end{cases}$$

and

$$\hat{\nu}_{\lambda_i, \lambda_i}(\psi, \hat{\lambda}_{i\psi}) = \begin{cases} \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2 & \text{if } \psi = 0 \\ 2 \sum_{j=1}^m (y_{ij}^2 - \tilde{y}_i) / (\pi \sqrt{\tilde{y}_i}) & \text{if } \psi = 1. \end{cases}$$

Table 1 gives estimated probabilities of correct selection under the stratified exponential model (Exp) and under the stratified half-normal model (Hn), based on 10,000 samples. Different stratum sizes m are considered ranging from 3 to 20 with a number q of strata increasing from 1 to 20 (the case $m = 3$ and $q = 1$ has been omitted).

For all values of m and q , the marginal likelihood gives the highest sum of estimated probabilities of correct selection. This is in line with the Neyman-Pearson optimality property among invariant tests of the likelihood ratio test $L_I(1)/L_I(0)$ with critical value equal to 1 (see e.g. Severini, 2000, Section 3.2). Estimated probabilities of correct selection for $\ell_L(\psi)$ are quite close to those for $\ell_I(\psi)$ for $m \geq 5$.

When $q = 1$, i.e. in the usual scenario of unstratified samples, all the procedures behave similarly as m increases, with some unbalance between the probabilities of correct selection under the two models when $m = 5$ for $\ell_{AIC}(\psi)$ and $\ell_{TIC}(\psi)$.

With small m ($m = 3, 5$), the traditional model selection criteria $\ell_{AIC}(\psi)$ and $\ell_{TIC}(\psi)$ show severe unbalance as q increases, and a form of Neyman-Scott phenomenon emerges. For $m = 20$ all the criteria become comparable with some residual unbalance for $\ell_{AIC}(\psi)$ and $\ell_{TIC}(\psi)$.

Example 2: Selection between stratified lognormal and inverse Gaussian models.

Suppose that, under the first model, where $\psi = 0$, Y_{ij} has a lognormal distribution with mean λ_i and variance $(k\lambda_i)^2$, i.e.

$$p(y_{ij}; 0, \lambda_i) = \frac{1}{\sqrt{2\pi c y_{ij}}} \exp \left\{ -\frac{1}{2c} (\log y_{ij} + 0.5c - \log \lambda_i)^2 \right\},$$

with $c = \log(k^2 + 1)$ and k a known constant. Under the second model, $\psi = 1$, Y_{ij} has an inverse Gaussian distribution with mean λ_i and variance λ_i^2 , i.e.

$$p(y_{ij}; 1, \lambda_i) = \frac{\sqrt{\lambda_i}}{\sqrt{2\pi y_{ij}^3}} \exp \left\{ -\frac{(y_{ij} - \lambda_i)^2}{2\lambda_i y_{ij}} \right\}.$$

Let $y_i^\dagger = m^{-1} \sum_{j=1}^m \log y_{ij}$ and $\tilde{y}_i = m^{-1} \sum_{j=1}^m y_{ij}^{-1}$, $i = 1, \dots, q$. We have

$$\hat{\lambda}_{i\psi} = \hat{\lambda}_{i0} = \exp \left(\sum_{j=1}^m (\log y_{ij} + 0.5c) / m \right)$$

when $\psi = 0$, and

$$\hat{\lambda}_{i\psi} = \hat{\lambda}_{i1} = (1 + \sqrt{1 + 4\tilde{y}_i y_i^\dagger}) / (2\tilde{y}_i)$$

when $\psi = 1$. The profile loglikelihood is

$$\ell_P(\psi) = \begin{cases} -\frac{mq}{2} \log(2\pi c) - \sum_{i,j} \log y_{ij} - \frac{1}{2c} \sum_{i,j} (\log y_{ij} - y_i^\dagger)^2 & \text{if } \psi = 0 \\ q(m + \log 2 - \frac{m}{2} \log(2\pi)) - \frac{3}{2} \sum_{i,j} \log y_{ij} + \frac{m}{2} \sum_{i=1}^q \left(\log \hat{\lambda}_{i1} - \frac{\bar{y}_i}{\hat{\lambda}_{i1}} - \tilde{y}_i \hat{\lambda}_{i1} \right) & \text{if } \psi = 1. \end{cases}$$

As in the previous example, $\ell_P(\psi)$ is equivalent to $\ell_{AIC}(\psi)$.

The marginal loglikelihood based on (4) is

$$\ell_I(\psi) = \begin{cases} -\frac{1}{2c} \sum_{i,j} (\log y_{ij} - y_i^\dagger)^2 - \frac{q(m-1)}{2} \log(2\pi c) - \frac{q}{2} \log m \sum_{i,j} \log y_{ij} & \text{if } \psi = 0 \\ q(m + \log 2 - \frac{m}{2} \log(2\pi)) - \frac{3}{2} \sum_{i,j} \log y_{ij} + \frac{m}{4} \sum_{i=1}^q \log \frac{\bar{y}_i}{\tilde{y}_i} + \sum_{i=1}^q \log a_i & \text{if } \psi = 1, \end{cases}$$

with $a_i = K_{\frac{m}{2}}(m\sqrt{\bar{y}_i \tilde{y}_i})$, where $K_\nu(x)$ is the modified Bessel function of the second kind (Abramowitz and Stegun, 1972, Section 9.6).

For computing $\ell_L(\psi)$ and $\ell_{TIC}(\psi)$, we need

$$j_{\lambda_i \lambda_i}(\psi, \hat{\lambda}_{i\psi}) = \begin{cases} \frac{m}{c \hat{\lambda}_{i0}^2} & \text{if } \psi = 0 \\ \frac{m}{2 \hat{\lambda}_{i1}^2} + \frac{m \bar{y}_i}{\hat{\lambda}_{i1}^3} & \text{if } \psi = 1 \end{cases}$$

and

$$\hat{\nu}_{\lambda_i, \lambda_i}(\psi, \hat{\lambda}_{i\psi}) = \begin{cases} \frac{1}{c^2 \hat{\lambda}_{i0}^2} \sum_{j=1}^m (\log y_{ij} - y_i^\dagger)^2 & \text{if } \psi = 0 \\ \frac{1}{4 \hat{\lambda}_{i1}^2} \sum_{j=1}^m \left(1 + \frac{y_{ij}}{\hat{\lambda}_{i1}} - \frac{\hat{\lambda}_{i1}}{y_{ij}} \right)^2 & \text{if } \psi = 1. \end{cases}$$

Table 2 gives estimated probabilities of correct selection, based on 10,000 samples and using $k = 1.2$. Different stratum sizes m are considered ranging from 3 to 20 with a number q of strata increasing from 1 to 20 (the case $m = 3$ and $q = 1$ has been omitted).

As expected, the marginal loglikelihood $\ell_I(\psi)$ gives the highest sum of estimated probabilities of correct selection. As the number of strata increases with $m \geq 5$, results for the Laplace approximation $\ell_L(\psi)$ become comparable with those for the optimal procedure based on $\ell_I(\psi)$. On the other hand, the traditional model selection criteria $\ell_{AIC}(\psi)$ and $\ell_{TIC}(\psi)$ show severe unbalance for with small m , especially as q increases.

4 Discussion

Findings of the paper are limited to stratified group families. Related results for scale and regression models are in Pace *et al.* (2006). Preliminary simulations indicate that Neyman-Scott phenomena occur when discriminating between separate stratified exponential models as well. The practical implication is that, if correct selection probabilities are considered important, model selection should be based on exact or accurate elimination of nuisance parameters. In group models this requirement is easily met via the marginal likelihood and its Laplace expansion. No analogous reduction seems to be available for exponential or more general models and how to introduce inferentially sound criteria is an open problem. These new criteria should extend to a discrete parameter of interest, in the presence of many nuisance parameters, the nuisance parameter elimination carried out by the modified profile likelihood in modern likelihood theory (see e.g. Severini, 2000, Chapter 9 and Sartori, 2003).

Loglikelihood True model	$\ell_{AIC}(\psi)$		$\ell_{TIC}(\psi)$		$\ell_I(\psi)$		$\ell_L(\psi)$	
	Exp	Hn	Exp	Hn	Exp	Hn	Exp	Hn
$m = 3, q = 3$	0.12	0.97	0.00	1.00	0.59	0.65	0.62	0.60
$m = 3, q = 10$	0.01	0.99	0.00	1.00	0.68	0.71	0.77	0.59
$m = 3, q = 20$	0.00	1.00	0.00	1.00	0.76	0.78	0.86	0.68
$m = 5, q = 1$	0.34	0.83	0.31	0.85	0.55	0.65	0.55	0.65
$m = 5, q = 3$	0.29	0.94	0.23	0.95	0.65	0.71	0.67	0.69
$m = 5, q = 10$	0.19	0.99	0.09	1.00	0.80	0.82	0.80	0.81
$m = 5, q = 20$	0.10	1.00	0.03	1.00	0.88	0.90	0.88	0.88
$m = 10, q = 1$	0.50	0.83	0.57	0.77	0.63	0.71	0.63	0.71
$m = 10, q = 3$	0.55	0.94	0.67	0.87	0.76	0.80	0.79	0.79
$m = 10, q = 10$	0.62	1.00	0.83	0.97	0.92	0.93	0.92	0.91
$m = 10, q = 20$	0.69	1.00	0.92	1.00	0.98	0.98	0.98	0.98
$m = 20, q = 1$	0.64	0.85	0.71	0.78	0.72	0.78	0.73	0.78
$m = 20, q = 3$	0.78	0.95	0.88	0.89	0.88	0.90	0.88	0.90
$m = 20, q = 10$	0.93	0.99	0.98	0.98	0.98	0.99	0.98	0.99
$m = 20, q = 20$	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Table 1: Estimated probabilities of correct selection under stratified exponential (Exp) and half-normal (Hn) models, Example 1.

Loglikelihood True model	$\ell_{AIC}(\psi)$		$\ell_{TIC}(\psi)$		$\ell_I(\psi)$		$\ell_L(\psi)$	
	Ln	Ig	Ln	Ig	Ln	Ig	Ln	Ig
$m = 3, q = 3$	0.14	0.96	0.22	0.77	0.45	0.73	0.40	0.78
$m = 3, q = 10$	0.12	0.99	0.19	0.81	0.57	0.74	0.46	0.82
$m = 3, q = 20$	0.08	0.99	0.17	0.84	0.65	0.75	0.53	0.84
$m = 5, q = 1$	0.24	0.86	0.30	0.75	0.38	0.75	0.34	0.76
$m = 5, q = 3$	0.28	0.92	0.36	0.74	0.51	0.74	0.49	0.76
$m = 5, q = 10$	0.28	0.98	0.40	0.79	0.67	0.77	0.62	0.81
$m = 5, q = 20$	0.25	0.99	0.42	0.82	0.77	0.82	0.71	0.87
$m = 10, q = 1$	0.39	0.84	0.45	0.72	0.50	0.75	0.49	0.75
$m = 10, q = 3$	0.46	0.89	0.56	0.73	0.61	0.77	0.61	0.77
$m = 10, q = 10$	0.57	0.98	0.72	0.82	0.80	0.88	0.79	0.89
$m = 10, q = 20$	0.67	0.99	0.83	0.89	0.90	0.95	0.89	0.96
$m = 20, q = 1$	0.51	0.80	0.56	0.72	0.58	0.75	0.58	0.75
$m = 20, q = 3$	0.65	0.91	0.74	0.79	0.75	0.83	0.75	0.83
$m = 20, q = 10$	0.80	0.98	0.90	0.90	0.91	0.95	0.91	0.95
$m = 20, q = 20$	0.93	0.99	0.97	0.96	0.98	0.99	0.98	0.99

Table 2: Estimated probabilities of correct selection under stratified lognormal (Ln) with $k = 1.2$ and inverse Gaussian (Ig) models, Example 2.

References

- Abramowitz, M., Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York.
- Akaike, H. (1973). Information theory and the maximum likelihood principle. *2nd International Symposium on Information Theory*, Akademiai Kiado, Budapest.
- Barndorff-Nielsen, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343-365.
- Burnham, K.P., Anderson, D.R. (2002). *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*. Second edition, Springer, New York.
- Ducharme, G.R., Frichot, B. (2003). Quasi most powerful invariant goodness-of-fit tests. *Scand. J. Stat.*, **30**, 399-414.
- Neyman, J., Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1-32.
- Pace, L., Salvan, A., Ventura, L. (2006). Likelihood based discrimination between separate scale and regression models. *J. Stat. Plan. Inf.*, 3539–3553.
- Quesenberry, C.P. (1985). Model construction: Selection of distributions. In *Encyclopedia of Statistical Sciences*, Vol. 5, 583–589.
- Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika*, **90**, 533–549.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464
- Severini, T. (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.
- Strupczewski, W.G., Mitosek, H.T., Kochanek, K., Singh, V.P., Weglarczyk, S. (2006). Probability of correct selection from lognormal and convective diffusion models based on the likelihood ratio. *Stoch. Env. Res. Risk A.*, **20**, 152–163.
- Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Mathematical Sciences*, **153**, 12–18. (in Japanese)

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: <http://wp.stat.unipd.it>

