Ann. Inst. Statist. Math.
Vol. 55, No. 1, 187–195 (2003)

A NOTE ON LIKELIHOOD ASYMPTOTICS IN NORMAL LINEAR REGRESSION

N. SARTORI

Department of Statistics, University of Padova, via C. Battisti 241-243, I-35121 Padova, Italy,
e-mail: sartori@stat.unipd.it

(Received August 15, 2000; revised February 26, 2002)

Abstract. Higher-order likelihood methods often give very accurate results. A way to evaluate accuracy is the comparison of the solutions with the exact ones of the classical theory, when these exist. To this end, we consider inference for a scalar regression parameter in the normal regression setting. In particular, we compare confidence intervals computed from the likelihood and its higher-order modifications with the ones based on the Student t distribution. It is shown that higher-order likelihood methods give accurate approximations to exact results.

Key words and phrases: Adjusted profile likelihoods, confidence interval, higher-order asymptotics, modified directed likelihood, nuisance parameter, orthogonal parameterization.

1. Introduction

Standard first-order methods for inference on individual components of a multi-dimensional parameter can be seriously inaccurate in the presence of nuisance parameters. The generally best first-order method is based on the χ_1^2 approximation to the distribution of the usual loglikelihood ratio statistic or on the normal approximation to the signed squared root of the loglikelihood ratio, also called the directed likelihood. These can be thought of as the corresponding statistics computed from the profile likelihood, acting as though it were the likelihood for a one parameter model. From this viewpoint, the main reason for inaccuracy is that the profile likelihood does not take into account the effects of fitting nuisance parameters. In particular, the expected value of the profile score is in general of order $O(1)$ (McCullagh and Tibshirani (1990)).

Higher-order asymptotic results can be used both to construct adjusted profile likelihoods and to improve the accuracy of the asymptotic distribution of likelihood based tests. Proposals of the first kind have been made by Barndorff-Nielsen (1983, 1994), Cox and Reid (1987), McCullagh and Tibshirani (1990), among others. These adjusted profile likelihoods have a score bias of order $O(n^{-1})$, where n is the sample size. More generally, typically adjusted profile likelihoods incorporate a traslation and a curvature effect on the profile likelihood. For the example considered in this paper, we just have the curvature effect. This may be considerable when there are many nuisance parameters. With the second aim, a statistic for inference on a scalar parameter of interest with high accuracy is the modified directed likelihood of Barndorff-Nielsen (1986, 1991). In particular, under the null hypothesis, the modified directed likelihood has a standard normal distribution to third order.

The modified directed likelihood requires the calculation of some sample space derivatives, i.e. derivatives with respect to components of the maximum likelihood estimate. This means that one needs to write the minimal sufficient statistic as a one-to-one function of the maximum likelihood estimate and a suitable ancillary statistic, either exactly or approximately. Outside full rank exponential models and transformation models this may be difficult. Sample space derivatives are required also in the computation of the modified profile likelihood (Barndorff-Nielsen (1983)). Recently, approximations for the sample space derivatives have been proposed in the literature. Using these approximations, one can obtain versions of the modified directed likelihood (and of the modified profile likelihood) that still have good accuracy properties. See §7.5 (and §9.5) of Severini (2000) for a recent review.

Simulation results (DiCiccio and Martin (1993); DiCiccio and Stern (1994); Sartori *et al.* (1999)) have shown that inference based on the modified profile likelihood and, in particular, on the modified directed likelihood is quite accurate, even in the presence of many nuisance parameters.

The approach here is slightly different. We want to investigate how results of higher-order likelihood asymptotics yield highly accurate approximation to an exact optimal solution, when this exists. In particular, we consider the classical problem of inference in the normal linear model and we compare confidence intervals based on Student t distribution with those based on likelihood and its modifications. Third-order methods in this context have been investigated through simulations in Fraser *et al.* (1999*b*).

Section 2 gives some notation and background on likelihood asymptotics. In particular, a two step modified directed likelihood is proposed. Section 3 deals with inference on a scalar regression coefficient in a normal linear model, treating the other parameters as nuisance parameters. Finally, a brief discussion is given in Section 4.

2. Notation and preliminaries

Consider a parametric statistical model with probability density function $p(y;\theta)$. The parameter θ has dimension d and is partitioned as (ψ, λ) into a scalar parameter of interest ψ and a nuisance parameter λ, of dimension $d-1$.

Suppose that we can write the data y as a one-to-one function of $(\hat{\theta}, a)$, where $\hat{\theta}$ is the maximum likelihood estimator of θ and a is an ancillary statistic, either exactly or approximately. The vector of first-order derivatives of the loglikelihood function $l(\theta; \hat{\theta}, a)$ with respect to a subset ρ of components of θ, such as λ or ψ, will be denoted by $l_\rho(\theta)$, whereas the vector of sample space derivatives of the likelihood, with respect to $\hat{\nu}$ consisting of components of $\hat{\theta}$, will be denoted by $l_{;\hat{\nu}}(\theta)$. For higher-order derivatives we will use symbols such as $l_{\rho;\hat{\nu}}(\theta)$ and $l_{\rho\nu}(\theta)$. Similarly, $j_{\rho\nu}(\theta)$ will denote blocks of the observed information function $j(\theta)$.

The profile loglikelihood $l(\hat{\theta}_\psi)$ will be denoted by $l_P(\psi)$, where $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$ and with $\hat{\lambda}_\psi$ denoting the constrained maximum likelihood estimator of λ for a given value of ψ. The signed square root of the likelihood ratio statistic, also known as signed likelihood root, is

$$r(\psi) = \text{sgn}(\hat{\psi} - \psi)\{l_P(\hat{\psi}) - l_P(\psi)\}^{1/2}$$

and will be called the directed likelihood.

2.1 *Adjusted profile likelihoods*

The modified profile loglikelihood (Barndorff-Nielsen (1983)) is

$$l_M(\psi) = l_P(\psi) + \log C(\psi),$$

with

$$C(\psi) = \frac{\{|j_{\lambda\lambda}(\hat{\theta})||j_{\lambda\lambda}(\hat{\theta}_\psi)|\}^{1/2}}{|l_{\lambda;\hat{\lambda}}(\hat{\theta}_\psi)|},$$

where the data-dependent factor $j_{\lambda\lambda}(\hat{\theta})$ is introduced to clarify the relation between $l_M(\psi)$ and the modified directed likelihood, introduced in the next section. The modified profile likelihood has a central role among adjusted profile likelihoods, because of its desirable properties. First of all, it approximates a conditional or a marginal likelihood, when either exists. Moreover, it is invariant to interest respecting reparameterizations and satisfies to second order the first two Bartlett relations (DiCiccio *et al.* (1996)).

If ψ and λ are orthogonal, that means $i_{\psi\lambda} = 0$, the modified profile loglikelihood is second order equivalent to the approximate conditional loglikelihood of Cox and Reid (1987)

$$l_{AC}(\psi) = l_P(\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\hat{\theta}_\psi)|.$$

The approximate conditional likelihood is simpler to compute than the modified profile likelihood but it is not invariant to interest respecting reparameterizations.

We will use $r_M(\psi)$ and $r_{AC}(\psi)$ to denote the directed likelihood computed from the approximate conditional likelihood and from the modified profile likelihood, respectively.

2.2 *Modified directed likelihood*

The modified directed likelihood (Barndorff-Nielsen (1986, 1991)) is

$$r^*(\psi) = r(\psi) + \frac{1}{r(\psi)}\log\frac{u(\psi)}{r(\psi)},$$

with

$$u(\psi) = |l_{;\hat{\theta}}(\hat{\theta}) - l_{;\hat{\theta}}(\hat{\theta}_\psi) l_{\lambda;\hat{\theta}}(\hat{\theta}_\psi)| / \{|j_{\lambda\lambda}(\hat{\theta}_\psi)||j(\hat{\theta})|\}^{1/2}.$$

A different version of $u(\psi)$ is given by formula (2.7) of Fraser *et al.* (1999*a*). However, in the normal distribution it coincides with the Barndorff-Nielsen formula. For this reason we do not give it in detail here, though it might be of more general application.

Pierce and Peters (1992) point out that in exponential families the adjustment to $r(\psi)$ giving $r^*(\psi)$ can be decomposed as the sum of two terms, and Barndorff-Nielsen and Cox ((1994), § 6.6.4) note the general existence of such a decomposition. In particular $r^*(\psi)$ can be expressed as

$$r^*(\psi) = r(\psi) + NP + INF,$$

where NP is the nuisance parameters adjustment

$$NP = -\frac{1}{r(\psi)}\log C(\psi),$$

and INF is the information adjustment

$$INF = \frac{1}{r(\psi)} \log \frac{u_P(\psi)}{r(\psi)},$$

with

$$u_P(\psi) = j_P(\hat{\psi})^{-1/2} \frac{\partial}{\partial \hat{\psi}} \{l_P(\hat{\psi}) - l_P(\psi)\}. \tag{2.1}$$

Here, j_P is the profile observed information, and the derivatives with respect to $\hat{\psi}$ are calculated with $l_P(\hat{\psi}) - l_P(\psi)$ considered as a function of $\psi, \hat{\psi}, \hat{\lambda}_\psi$ and a. The empirical results in Pierce and Peters (1992) indicate that the NP adjustment is often considerable, and it may yield a more substantial effect than the INF adjustment, especially when the dimension of the nuisance parameter is large. Moreover, Sartori *et al.* (1999) show that $r_M(\psi) = r(\psi) + NP + O(n^{-1})$ and give simulation results which indicate that, as the number of nuisance parameters increases, inference based on $r_M(\psi)$ is very close to that based on $r^*(\psi)$.

2.3 *Two step modified directed likelihood*

Pierce and Peters (1992) suggest a two step procedure to third-order inference for a scalar canonical parameter in full exponential families. In the first step, an adjusted profile likelihood for the interest parameter is obtained. In the second step, the modified directed likelihood for this one parameter pseudo-likelihood is computed.

More generally, if we work from the modified profile likelihood and calculate the INF adjustment, as was done leading to (2.1), we obtain

$$r_M^*(\psi) = r_M(\psi) + \frac{1}{r_M(\psi)} \log \frac{u_M(\psi)}{r_M(\psi)}, \tag{2.2}$$

where

$$u_M(\psi) = j_M(\hat{\psi}_M)^{-1/2} \frac{\partial}{\partial \hat{\psi}_M} \{l_M(\hat{\psi}_M) - l_M(\psi)\} \tag{2.3}$$

and with j_M and $\hat{\psi}_M$ denoting respectively the observed information and the maximizer of $l_M(\psi)$. We can write (2.3) in the following form

$$u_M(\psi) = j_M(\hat{\psi}_M)^{-1/2} (\partial \hat{\psi}_M / \partial \hat{\psi})^{-1} \frac{\partial}{\partial \hat{\psi}} \{l_M(\hat{\psi}_M) - l_M(\psi)\},$$

where

$$(\partial \hat{\psi}_M / \partial \hat{\psi})^{-1} = j_M(\hat{\psi}_M)^{-1} (\partial^2 l_M(\psi) / \partial \psi \partial \hat{\psi})|_{\psi = \hat{\psi}_M} \tag{2.4}$$

is obtained differentiating the likelihood equation for $\hat{\psi}_M$ with respect to $\hat{\psi}$. In general, it can be analytically cumbersome to compute (2.3) and a different version (simpler to compute) would be preferable. An alternative way is the one proposed in DiCiccio *et al.* (2001), which makes use of simulation. Anyway, in the present context (2.4) is equal to 1, because $\hat{\psi}_M = \hat{\psi}$, and it is possible to express $u_M(\psi)$ as a simple modification of $u_P(\psi)$.

All directed likelihoods are approximate pivotal quantities, which allow us to obtain p-values and confidence limits for ψ. However, while $r(\psi)$, $r_{AC}(\psi)$ and $r_M(\psi)$ are standard normal only to first-order, $r^*(\psi)$ and $r^*_M(\psi)$ are standard normal with third-order accuracy. Note that $r(\psi)$, $r_M(\psi)$, $r^*(\psi)$ and $r^*_M(\psi)$ are invariant under interest respecting reparameterizations, while $r_{AC}(\psi)$ is not.

3. Inference on a scalar regression parameter

Consider n random variables $Y_1, \dots, Y_n$ following a linear regression model $Y_i = \mu_i + \sigma\varepsilon_i$, where $\mu_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ $(i = 1, \dots, n)$ and $\varepsilon_1, \dots, \varepsilon_n$ are independent standard normal variables. The vectors $(x_{i1}, \dots, x_{ip})$ of covariates values are assumed to be known and the vector $\beta = (\beta_1, \dots, \beta_p)$ is the vector of regression parameters to be estimated. Using matrix notation, we can write $Y = X\beta + \sigma\epsilon$ where $Y = (Y_1, \dots, Y_n)$, $X = (x_{ij})$ is an $n \times p$ matrix of rank p $(n \geq p)$, called the regression matrix, and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$.

The likelihood function for $\theta = (\beta, \sigma^2)$ is

$$l(\theta; y) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sigma^{-2}(Y - X\beta)'(Y - X\beta)$$

and the maximum likelihood estimate is $\hat{\theta} = (\hat{\beta}, \hat{\sigma}^2)$ with

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \hat{\sigma}^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})/n.$$

As is well known, this is a full rank exponential family, so the maximum likelihood estimate is also sufficient.

Let a scalar component of the regression parameter β be the parameter of interest. Without loss of generality, we can consider $\psi = \beta_p$, i.e. the regression coefficient of the p-th explanatory variable. The nuisance parameter is $\lambda = (\gamma, \sigma^2)$, where $\gamma = (\beta_1, \dots, \beta_{p-1})$. An exact $1 - \alpha$ confidence interval for ψ is given by

$$(\hat{\psi} - t_{n-p,\alpha/2}s\sqrt{v_{pp}}, \hat{\psi} + t_{n-p,\alpha/2}s\sqrt{v_{pp}}),$$

where $s^2 = n\hat{\sigma}^2/(n-p)$, $\hat{\psi} = \hat{\beta}_p$, v_{pp} is the entry at position (p,p) of $V = (X'X)^{-1}$ and $t_{n-p,\alpha/2}$ is the upper quantile of level $\alpha/2$ of the t distribution with $n - p$ degrees of freedom.

The profile loglikelihood function for ψ is

$$l_P(\psi) = l(\psi, \hat{\gamma}_\psi, \hat{\sigma}^2_\psi) = -\frac{n}{2}\log\hat{\sigma}^2_\psi,$$

where $\hat{\sigma}^2_\psi = \hat{\sigma}^2 + (nv_{pp})^{-1}(\hat{\psi} - \psi)^2$, $\hat{\gamma}_\psi = \hat{\gamma} + (X_1'X_1)^{-1}X_1'x_p(\hat{\psi} - \psi)$ and where we have considered the partition $X = [X_1, x_p]$ of the regression matrix. In particular, X_1 is the $n \times (p-1)$ matrix containing the values of the first $(p-1)$ explanatory variables and x_p is the p-th explanatory variable. We can use the directed likelihood to construct a $1 - \alpha$ confidence interval for ψ of the form $|r(\psi)| < z_{\alpha/2}$, with

$$r(\psi) = \text{sgn}(\hat{\psi} - \psi)\left(n\log\frac{\hat{\sigma}^2_\psi}{\hat{\sigma}^2}\right)^{1/2} = \text{sgn}(t)\left\{n\log\left(1 + \frac{t^2}{n-p}\right)\right\}^{1/2},$$

where $t = (\hat{\psi} - \psi)/(sv_{pp}^{1/2})$ and $z_{\alpha/2}$ is the $\alpha/2$ upper quantile of the standard normal distribution. The interval can be reexpressed in the form

$$(\hat{\psi} - c_n s\sqrt{v_{pp}}, \hat{\psi} + c_n s\sqrt{v_{pp}}), \tag{3.1}$$

with $c_n = \sqrt{n-p}\{\exp(z_{\alpha/2}^2/n) - 1\}^{1/2}$.

The modified profile likelihood for ψ is

$$l_M(\psi) = -\frac{n-p-1}{2}\log\hat{\sigma}_\psi^2 = \frac{n-p-1}{n} l_P(\psi)$$

and, using its directed likelihood, gives a $1-\alpha$ confidence interval of the form

$$(\hat{\psi} - c_{n,2} s\sqrt{v_{pp}}, \hat{\psi} + c_{n,2} s\sqrt{v_{pp}}), \tag{3.2}$$

with $c_{n,2} = \sqrt{n-p}[\exp\{z_{\alpha/2}^2/(n-p-1)\} - 1]^{1/2}$. Note that (3.1) and (3.2) are standard likelihood drop intervals obtained from $l_P(\psi)$ and $l_M(\psi)$, respectively.

Consider a Cornish-Fisher expansion for the quantiles of a Student t distribution with $n-p$ degrees of freedom

$$t_{n-p,\alpha/2} = z_{\alpha/2} + \frac{g(z_{\alpha/2})}{4(n-p)} + O\{(n-p)^{-2}\}, \tag{3.3}$$

where $g(x) = x + x^3$. Expanding c_n and $c_{n,2}$ and comparing these expansions with (3.3) we have

$$c_n = t_{n-p,\alpha/2} - \frac{2p+1}{4(n-p)} z_{\alpha/2} + O\{(n-p)^{-2}\}$$
$$c_{n,2} = t_{n-p,\alpha/2} + \frac{z_{\alpha/2}}{4(n-p)} + O\{(n-p)^{-2}\}.$$

Note that the error in the approximation of the Student quantile is in both cases of order $O\{(n-p)^{-1}\}$. If the number of nuisance parameter is considerable with respect to the degrees of freedom, using the profile likelihood we have an error of order $O\{p/(n-p)\}$. On the other hand, for the modified profile likelihood the error is still of order $O\{(n-p)^{-1}\}$.

We can also use the modified directed likelihood. The needed quantities are $r(\psi)$,

$$C(\psi) = \{1 + t^2/(n-p)\}^{(p+1)/2} \qquad u_P(\psi) = \sqrt{n/(n-p)}t.$$

A $1-\alpha$ confidence interval based on $r^*(\psi)$ can be written in the asymptotic equivalent form $(\hat{\psi} - c_n^* s\sqrt{v_{pp}}, \hat{\psi} + c_n^* s\sqrt{v_{pp}})$, where

$$c_n^* = \operatorname{sgn}(z_{\alpha/2})\sqrt{n-p}\left[\exp\left\{z_{\alpha/2}^2\left(1 + \frac{2p+1}{4}n^{-1}\right)^2 n^{-1}\right\} - 1\right]^{1/2}.$$

If we expand c_n^* we have

$$\begin{aligned} c_n^* &= z_{\alpha/2}\left[1 + \frac{1}{4(n-p)}\{1 + z_{\alpha/2}^2\} + O\{(n-p)^{-2}\}\right] \\ &= t_{n-p,\alpha/2} + O\{(n-p)^{-2}\}, \end{aligned}$$

i.e. c_n^* is a highly accurate approximation of $t_{n-p,\alpha/2}$.

The two step modified directed likelihood is given by (2.2), with

$$u_M(\psi) = \sqrt{(n-p-1)/(n-p)}t.$$

As for the modified directed likelihood, we can obtain a confidence interval for ψ, using $r_M^*(\psi)$. It can be shown that the accuracy in the approximation to the exact interval is the same as that of $r^*(\psi)$.

Consider now the regression matrix $X = [X_1, x_p]$ as block orthogonal, i.e. $X_1' x_p = 0$. In this case the parameter of interest ψ is orthogonal to the nuisance parameter (γ, σ^2) and we can compute also the approximate conditional likelihood of Cox and Reid. Since $l_{AC}(\psi)$ is dependent on the chosen orthogonal parameterization, we wonder if there is an optimal parameterization among those of the form

$$(\psi, \gamma, \sigma^j). \tag{3.4}$$

When $j = 0$ we may think to use the parameterization $(\psi, \log\sigma)$, which gives $l_{AC}(\psi) = l_P(\psi)$. In particular, we choose the value of j which gives an approximate conditional likelihood, whose confidence intervals agree to a higher-order with the exact ones. The directed likelihood calculated from the approximate conditional likelihood in the parameterization (3.4) is

$$r_{AC}^{(j)}(\psi) = \left\{(n-p-j+1)\log\left(1 + \frac{t^2}{n-p}\right)\right\}^{1/2},$$

and gives a $1-\alpha$ approximate confidence interval for ψ of the form $(\hat{\psi} - c_{n,j}s\sqrt{v_{pp}}, \hat{\psi} + c_{n,j}s\sqrt{v_{pp}})$, with

$$c_{n,j} = \sqrt{n-p}[\exp\{z_{\alpha/2}^2/(n-p-j+1)\} - 1]^{1/2}.$$

An expansion for $c_{n,j}$ gives

$$\begin{aligned} c_{n,j} &= z_{\alpha/2}\left[1 + \frac{1}{4(n-p)}\{2(j-1) + z_{\alpha/2}^2\} + O\{(n-p)^{-2}\}\right] \\ &= t_{n-p,\alpha/2} + \frac{2j-3}{4(n-p)} z_{\alpha/2} + O\{(n-p)^{-2}\}. \end{aligned}$$

If we choose $j = 3/2$, we obtain an highly accurate approximation to the exact result. In particular, the order of approximation is the same as the one achievable with modified directed likelihoods. This suggests a method for choosing among orthogonal parameterizations when an exact solution is not available. In fact, it could be possible to choose the orthogonal parameterization such that confidence intervals based on $l_{AC}(\psi)$ agree with those based on $r^*(\psi)$ to third order.

Note that all likelihoods quantities considered in this section can be written as a function of t, $n-p$ and p. This allows a straightforward comparison with exact results. Tables 1–3 give coverage probabilities of the exact Student interval transformed in the directed likelihood scale and evaluated with the standard normal distribution, for a case with $n = 10$ and various numbers of explanatory variables.

Both $r(\psi)$ and $r_M(\psi)$ are not very accurate, but $r(\psi)$ is more affected by high values of p. On the contrary, $r_{AC}^{(3/2)}(\psi)$, $r^*(\psi)$ and $r_M^*(\psi)$ give good approximations for the exact results, even though, as p increases, $r^*(\psi)$ seems to loose its usual "hyper accuracy". An explanation of this is given in the next section.

Table 1. Coverage probabilities of the 99% Student's interval, transformed in the r scale and evaluated with the standard normal distribution; $n = 10$ and $p = 2, 3, 4, 5$.

p	r	r_M	$r_{AC}^{(3/2)}$	r^*	r_M^*
2	99.70	98.69	98.97	99.06	99.00
3	99.85	98.62	98.97	99.15	99.00
4	99.94	98.53	98.95	99.27	99.00
5	99.99	98.39	98.93	99.44	99.00

Table 2. Coverage probabilities of the 95% Student's interval, transformed in the r scale and evaluated with the standard normal distribution; $n = 10$ and $p = 2, 3, 4, 5$.

p	r	r_M	$r_{AC}^{(3/2)}$	r^*	r_M^*
2	97.60	94.11	94.94	95.20	94.99
3	98.46	93.95	94.92	95.47	94.99
4	99.15	93.71	94.89	95.89	94.98
5	99.63	93.36	94.84	96.51	94.98

Table 3. Coverage probabilities of the 90% Student's interval, transformed in the r scale and evaluated with the standard normal distribution; $n = 10$ and $p = 2, 3, 4, 5$.

p	r	r_M	$r_{AC}^{(3/2)}$	r^*	r_M^*
2	94.20	88.72	89.93	90.31	89.98
3	95.81	88.50	89.91	90.71	89.98
4	97.29	88.18	89.87	91.36	89.97
5	98.52	87.69	89.81	92.34	89.97

4. Discussion

Higher-order likelihood methods give very accurate approximation for exact optimal results in the normal distribution. The best solutions are obtained with the approximate conditional likelihood of Cox and Reid and with the modified directed likelihoods, in particular with the one computed from the modified profile likelihood. However, in the first case, the results are strongly dependent on the *ad hoc* "optimal" parameterization. On the contrary, $r_M^*(\psi)$ and $r^*(\psi)$, are invariant under interest respecting reparameterizations and they do not need the regression matrix to be block orthogonal.

We note that $r_M(\psi)$, $r_{AC}^{(3/2)}(\psi)$ and $r_M^*(\psi)$ can be expressed as functions only of t and the degrees of freedom $n - p$, while $r(\psi)$ and $r^*(\psi)$ depend on t, $n - p$ and also p. This means that, using $r_M(\psi)$, $r_{AC}^{(3/2)}(\psi)$ or $r_M^*(\psi)$, the approximation for the Student distribution with $n-p$ degrees of freedom is "stable", i.e. we have the same approximation with (n, p) or with $(n + k, p + k)$, where k is an integer ($k \geq 1 - p$). The same is not true for $r(\psi)$ and $r^*(\psi)$. This explains why in Tables 1–3, as p increases, changes in the coverage probabilities are more evident in the second and in the fifth columns. This is even more evident if we compare $r^*(t)$ and $r_M^*(t)$ in two cases: (a) one with $n = 4$

and $p = 1$ and (b) one with $n = 7$ and $p = 4$. In both cases, we are considering approximations for a Student distribution with 3 degree of freedom. The 95% exact confidence interval has in both cases a correspondent $r_M^*(t)$ confidence interval of level 94.99%. Coverage probabilities of the corresponding $r^*(t)$ confidence intervals are (a) 95.37% and (b) 97.18%.

This last consideration might suggest that the two step procedure leading to $r_M^*(\psi)$ should be considered better than the direct calculation of $r^*(\psi)$. This is surely true when the number of nuisance parameter is not moderate. However, in general $r^*(\psi)$ gives very accurate results and, when they are not very accurate, they still are acceptable. Moreover, the calculation of $u_M(\psi)$ in the present context has been simplified for a particular feature of the normal case. In general models, to obtain $u_M(\psi)$ is not so straightforward and further work in this direction is needed.

Acknowledgements

The author would like to thank Luigi Pace, Nancy Reid and Alessandra Salvan for inspiration and useful comments.

References

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator, *Biometrika*, **70**, 343–65.

Barndorff-Nielsen, O. E. (1986). Inference of full and partial parameters based on the standardized signed log likelihood ratio, *Biometrika*, **73**, 307–322.

Barndorff-Nielsen, O. E. (1991). Modified signed log likelihood ratio, *Biometrika*, **78**, 557–563.

Barndorff-Nielsen, O. E. (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood, *J. Roy. Statist. Soc. Ser. B*, **56**, 125–140.

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

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion), *J. Roy. Statist. Soc. Ser. B*, **49**, 1–39.

DiCiccio, T. J. and Martin, M. A. (1993). Simple modifications for signed roots of likelihood ratio statistics, *J. Roy. Statist. Soc. Ser. B*, **55**, 305–316.

DiCiccio, T. J. and Stern, S. E. (1994). Constructing approximately standard normal pivots from signed roots of adjusted likelihood ratio statistics, *Scand. J. Statist.*, **21**, 447–460.

DiCiccio, T. J., Martin, M. A. and Stern, S. E. (2001). Simple and accurate one-sided inference from signed roots of likelihood ratios, *Canad. J. Statist.*, **29**, 67–76.

DiCiccio, T. J., Martin, M. A., Stern, S. E. and Young, G. A. (1996). Information bias and adjusted profile likelihoods, *J. Roy. Statist. Soc. Ser. B*, **58**, 189–203.

Fraser, D. A. S., Reid, N. and Wu, J. (1999*a*). A simple general formula for tail probabilities for frequentist and Bayesian inference, *Biometrika*, **86**, 249–264.

Fraser, D. A. S., Wong, A. and Wu, J. (1999*b*). Regression analysis, nonlinear or nonnormal: Simple and accurate *p*-values from likelihood analysis, *J. Amer. Statist. Soc.*, **94**, 1286–1295.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods, *J. Roy. Statist. Soc. Ser. B*, **52**, 325–344.

Pierce, D. A. and Peters, D. (1992). Practical use of higher-order asymptotics for multiparameter exponential families (with discussion), *J. Roy. Statist. Soc. Ser. B*, **54**, 701–737.

Sartori, N., Bellio, R., Salvan, A. and Pace, L. (1999). The directed modified profile likelihood in models with many nuisance parameters, *Biometrika*, **86**, 735–742.

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