A note on directed adjusted profile likelihoods

N. Sartori, A. Salvan, L. Pace

2000.11

Dipartimento di Scienze Statistiche
Università degli Studi
Via S. Francesco, 33
35121 Padova

Settembre 2000

A note on directed adjusted profile likelihoods

Nicola Sartori, Alessandra Salvan

Department of Statistics, University of Padova

Luigi Pace

Department of Statistics, University of Udine

Abstract

Several adjustments to the profile likelihood have been proposed in recent years, to take into proper account the effects of fitting nuisance parameters. Typically, these modifications reduce the score bias to order $O(n^{-1})$. Sometimes, the information bias is also reduced to $O(n^{-1})$. In some cases, adjusted profile likelihoods are higher-order approximations of suitable conditional or marginal target likelihoods. However, the adjustments seem to provide accurate inference also when an exact marginal or conditional target likelihood is not available. Here we consider adjusted profile likelihoods as approximations of a suitable general target likelihood. This is a likelihood for the parameter of interest defined in the least favourable direction. Literature supporting this choice of a target likelihood is reviewed. Attention is focused on a scalar parameter of interest. Some new results are obtained concerning the null and non null distribution of the directed likelihood calculated from an adjusted profile likelihood.

Some key words: Efficient score; Information bias; Local power; Nuisance parameter; Orthogeodesic model; Orthogonal parameter; Score bias; Tensor.

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 normal approximation to the signed square root of the usual likelihood ratio statistic. This can be thought of as the corresponding likelihood ratio statistic 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 fact, the score computed from the profile likelihood typically has bias of order $O(1)$. Several adjusted profile likelihoods, having a score bias of order $O(n^{-1})$, have been proposed in the literature, including the modified profile likelihood (Barndorff-Nielsen, 1983, 1994; see also Barndorff-Nielsen & Cox, 1994, § 8.2) and the proposals in Cox & Reid (1987), McCullagh & Tibshirani (1990). In particular, the modified profile likelihood is obtained as a higher-order approximation to a conditional or a marginal likelihood, when either exists. Moreover, among the proposed adjusted profile likelihoods with score bias of order $O(n^{-1})$, the modified profile likelihood has an information bias of order $O(n^{-1})$ as well. This is true even for the proposal of Stern (1997) and for the Cox and Reid approximate conditional likelihood calculated with a particular orthogonal parameterisation (DiCiccio et al., 1996).

Simulation results (DiCiccio & Martin, 1993; DiCiccio & Stern, 1994a; Sartori et al., 1999) show that inference based on the modified profile likelihood is quite accurate, even in the presence of many nuisance parameters. This is true in general and not only when a marginal or conditional likelihood exists. Severini (1998) provides a theoretical justification of this empirical findings, showing that the modified profile loglikelihood is a third-order local log-density function and is second-order locally informative. This paper describes a related argument. The approach here is to explicitly relate adjusted profile likelihoods to a suitable general target likelihood for inference about the parameter of interest.

The definition of a general target likelihood is not entirely obvious. The most naive idea is that it should be the likelihood calculated with the true value of the nuisance parameter. Under parameter orthogonality, this choice is supported by the arguments in Barndorff-Nielsen & Cox (1994), Mukerjee (1992) and DiCiccio & Efron (1996). Barndorff-Nielsen & Cox (1994, formula (3.50)) show that the profile likelihood ratio statistic is first-order equivalent to the likelihood ratio statistic when an orthogonal nuisance parameter is known. Moreover, Mukerjee (1992) shows that the likelihood ratio statistic based on Cox and Reid approximate conditional likelihood has, to second order, the same null and non-null distribution as the likelihood ratio statistic when an orthogonal nuisance parameter is known. The null distribution result is extended by Di Ciccio & Efron (1996) to any adjusted profile likelihood with score bias of order $O(n^{-1})$.

The results mentioned above suggest the likelihood with a known orthogonal nuisance parameter as a target likelihood. This choice seems particularly reasonable because of its relation with the concept of least favourable direction (cf., for example, Severini & Wong, 1992). The least favourable direction likelihood is invariant under interest respecting reparameterisations and is also an information preserving likelihood, in the sense that the corresponding expected information for the parameter of interest equals the information when the nuisance parameter is unknown.

Attention is focused here on a scalar parameter of interest. Following Mukerjee (1992) and DiCiccio and Efron (1996), we compare adjusted profile likelihoods and the target likelihood in terms of the null and non-null distribution of the corresponding directed likelihoods. As a first result, we extend the non-null distributional result of Mukerjee (1992) to any adjusted profile likelihood with score bias of order $O(n^{-1})$. We also find that, in general, the agreement of null distributions does not carry over to higher-order terms. Orthogeodesic models (Barndorff-Nielsen and Blæsild, 1993) are, however, one notable exception.

Section 2 gives some notation. The results of Mukerjee (1992) and of DiCiccio & Efron (1996, section 9) are briefly summarised in Section 3. In section 4, after recalling the relation between parameter orthogonality and least favourable direction, the non-null results of Mukerjee (1992) are extended to the general case and some higher order properties are investigated.

2 Notation

We consider a parametric statistical model with probability density function $p(y;\theta)$. The parameter θ has dimension d and is partitioned as (ψ, χ) into a scalar parameter of interest ψ and a nuisance parameter χ, of dimension $d-1$. Let $\hat{\theta}$ be the maximum likelihood estimator of θ and let $\tilde{\theta} = (\psi, \hat{\chi}_\psi)$, where $\hat{\chi}_\psi$ denotes the constrained maximum likelihood estimator of χ for a given value of ψ. The profile likelihood $L(\psi, \hat{\chi}_\psi)$ will be denoted by $L_P(\psi) = \exp\{l_P(\psi)\}$.

Suppose that we can write the minimal sufficient statistic as a one-to-one function of $(\hat{\theta}, a)$, where 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 χ or ψ, will be denoted by $l_\rho = l_\rho(\theta)$, whereas the vector of sample space derivatives of the loglikelihood, with respect to $\hat{\sigma}$ consisting of components of $\hat{\theta}$, will be denoted by $l_{;\hat{\sigma}} = l_{;\hat{\sigma}}(\theta)$. For higher-order derivatives we will use symbols such as $l_{\rho;\hat{\sigma}} = l_{\rho;\hat{\sigma}}(\theta)$ and $l_{\rho\sigma} = l_{\rho\sigma}(\theta)$. Similarly, $j_{\rho\sigma}$ and $i_{\rho\sigma}$ will denote blocks of the observed information j and of the expected information i, respectively. We will also let $H_{\rho\sigma} = l_{\rho\sigma} + i_{\rho\sigma}$. For null mean values of the derivatives of the loglikelihood we will write, for example, $\nu_{\rho\rho\sigma} = E_\theta\{l_{\rho\rho\sigma}(\theta)\}$ and $\nu_{\rho,\rho\sigma} = E_\theta\{l_\rho(\theta)\, l_{\rho\sigma}(\theta)\}$. For a generic function $g = g(\theta)$ we will denote by $g_{/\rho}$ the first order partial derivative with respect to ρ.

An adjusted profile loglikelihood for a possibly multidimensional ψ has the form

$$l_A(\psi) = l_P(\psi) + m(\psi)\,.$$

The approximate conditional loglikelihood of Cox & Reid (1987), $l_{AC}(\psi)$, requires an orthogonal parameterisation and has

$$m(\psi) = -\frac{1}{2}\log \mid j_{\chi\chi}(\tilde\theta) \mid\,.$$

The modified profile loglikelihood of Barndorff-Nielsen (1983), $l_{MP}(\psi)$, has

$$m(\psi) = -\frac{1}{2}\log \mid j_{\chi\chi}(\tilde\theta) \mid - \log \mid \hat\chi_{\psi/\hat\chi} \mid,$$

where

$$\mid \hat\chi_{\psi/\hat\chi} \mid = \frac{\mid l_{\chi;\hat\chi}(\tilde\theta) \mid}{\mid j_{\chi\chi}(\tilde\theta) \mid}\,.$$

Under random sampling of size n and orthogonal parameterisation, $l_{MP}(\psi) = l_{AC}(\psi) + O(n^{-1})$. This equation holds with error of order $O(n^{-3/2})$ if the orthogonal parameterisation satisfies one further condition given by Cox & Reid (1989).

We denote by m_ψ and $m_{\psi\psi}$ the first two derivatives of $m(\psi)$, and by μ_ψ and $\mu_{\psi\psi}$ their null mean values. An adjusted profile likelihood with score bias of order $O(n^{-1})$ satisfies the condition

$$\mu_\psi = \tau_\psi + O(n^{-1})\,, \tag{1}$$

where $-\tau_\psi$ is the term of order $O(1)$ of the mean value of the profile score function (see, for example, McCullagh & Tibshirani, 1990).

Let $l^\chi(\psi) = l(\psi, \chi)$ be the loglikelihood for ψ with χ fixed, and let $\hat\psi_\chi$ be the maximiser of $l^\chi(\psi)$. When ψ is a scalar and χ is known,

$$r_\chi(\psi) = \operatorname{sgn}(\hat\psi_\chi - \psi)\ \left[2\left\{l^\chi(\hat\psi_\chi) - l^\chi(\psi)\right\}\right]^{1/2}$$

is the signed square root of the likelihood ratio statistic for ψ, also called the directed likelihood for ψ. When χ is unknown, the directed likelihood for ψ is

$$r_P(\psi) = \text{sgn}(\hat{\psi} - \psi) \left[2\left\{l_P(\hat{\psi}) - l_P(\psi)\right\}\right]^{1/2}.$$

The maximiser of $l_P(\psi)$ is $\hat{\psi}$, so that $r_P(\psi)$ is the directed likelihood computed from $l_P(\psi)$. Similarly, we denote by $r_A(\psi)$ the directed adjusted profile likelihood, that is the directed likelihood computed from $l_A(\psi)$. In the following, we assume for simplicity that, as ψ, also χ is a scalar. Extensions of the results to a multidimensional nuisance parameter are straightforward.

3 Preliminary results

A standard expansion, see for instance Barndorff-Nielsen & Cox (1994), formula (5.41), shows that

$$r_\chi(\psi) = \frac{l_\psi}{\sqrt{i_{\psi\psi}}} + r_\chi^1(\psi) + O_p(n^{-1}), \tag{2}$$

where $r_\chi^1(\psi)$ is of order $O_p(n^{-1/2})$ and is given by

$$r_\chi^1(\psi) = \frac{1}{6i_{\psi\psi}^2}\left(\nu_{\psi\psi\psi}l_\psi + 3i_{\psi\psi}H_{\psi\psi}\right)\frac{l_\psi}{\sqrt{i_{\psi\psi}}}.$$

Therefore,

$$E_\theta(r_\chi(\psi)) = \frac{1}{6i_{\psi\psi}^{3/2}}(\nu_{\psi\psi\psi} + 3\nu_{\psi\psi,\psi}) + O(n^{-3/2}) = \frac{1}{6i_{\psi\psi}^{3/2}}t_3(\psi) + O(n^{-3/2}). \tag{3}$$

The quantity $t_3(\psi) = \nu_{\psi\psi\psi} + 3\nu_{\psi\psi,\psi} = -\nu_{\psi,\psi,\psi}$ appearing in (3) is a third order tensor under reparameterisations of ψ.

From now on in this section, let us assume that ψ and χ are orthogonal, namely that $i_{\psi\chi} = -E_\theta(l_{\psi\chi}) = 0$. Below, we summarise some of the main results from Mukerjee (1992). The directed likelihood for ψ with χ unknown has an expansion of the form

$$r_P(\psi) = r_\chi(\psi) + r_P^1(\psi) + O_p(n^{-1}),$$

where $r_P^1(\psi)$ is of order $O_p(n^{-1/2})$ and is given by

$$r_P^1(\psi) = \frac{1}{\sqrt{i_{\psi\psi} i_{\chi\chi}}} \left\{ H_{\psi\chi} l_\chi + \frac{1}{2i_{\chi\chi}} \nu_{\psi\chi\chi} l_\chi^2 - \frac{1}{2i_{\psi\psi}} (\nu_{\psi\psi\chi} + \nu_{\psi\psi,\chi}) l_\psi l_\chi \right\} .$$

Hence,

$$E_\theta(r_P^1(\psi)) = \frac{1}{\sqrt{i_{\psi\psi} i_{\chi\chi}}} \left(\nu_{\psi\chi,\chi} + \frac{1}{2} \nu_{\psi\chi\chi} \right) .$$

Moreover,

$$r_{AC}(\psi) = r_P(\psi) + \frac{1}{2\sqrt{i_{\psi\psi} i_{\chi\chi}}} \nu_{\psi\chi\chi} + O_p(n^{-1}) ,$$

where $r_{AC}(\psi)$ is the directed likelihood computed from $l_{AC}(\psi)$. Therefore,

$$\begin{aligned} E_\theta\left(r_{AC}(\psi)\right) &= E_\theta(r_\chi(\psi)) + \frac{1}{\sqrt{i_{\psi\psi} i_{\chi\chi}}} (\nu_{\psi\chi,\chi} + \nu_{\psi\chi\chi}) + O(n^{-3/2}) \\ &= E_\theta(r_\chi(\psi)) + O(n^{-3/2}) , \end{aligned}$$

where the last identity follows from the relation $\nu_{\psi\chi\chi} + \nu_{\psi\chi,\chi} = 0$, which is obtained differentiating with respect to χ the identity $E_\theta(l_{\psi\chi}) = 0$.

In addition,

$$\text{var}_\theta(r_{AC}(\psi)) = \text{var}_\theta(r_\chi(\psi)) + O(n^{-1}) = 1 + O(n^{-1}) ,$$

while third and higher order cumulants of both $r_{AC}(\psi)$ and $r_\chi(\psi)$ are of order $O(n^{-3/2})$. This implies that the Edgeworth expansions for the null distribution of $r_{AC}(\psi)$ and of $r_\chi(\psi)$ agree up to terms of order $O(n^{-1/2})$ included. Analogously, under local alternatives, the power functions of $r_{AC}(\psi)$ and of $r_\chi(\psi)$ coincide up to terms of order $O(n^{-1/2})$ included. Let $\psi = \psi_0 + \delta/\sqrt{n}$. The following expansions are reported for later use:

$$E_{\psi_0,\chi}(r_\chi(\psi)) = -\frac{\delta}{\sqrt{n}} i_{\psi\psi}^{1/2} + \frac{1}{6} i_{\psi\psi}^{-3/2} t_3(\psi) + \frac{\delta^2}{6n} i_{\psi\psi}^{-1/2} \nu_{\psi\psi\psi} + O(n^{-1}) , \quad (4)$$

$$\text{var}_{\psi_0,\chi}(r_\chi(\psi)) = 1 + \frac{\delta}{3\sqrt{n}} i_{\psi\psi}^{-1} t_3(\psi) + O(n^{-1}) , \quad (5)$$

while higher order cumulants of $r_\chi(\psi)$ under (ψ_0, χ) are of order $O(n^{-1})$.

DiCiccio & Efron (1996, section 9) extend the null distribution result to any adjusted profile likelihood invariant under interest respecting reparameterisations and with score bias of order $O(n^{-1})$.

4 Comparisons with inference from a target likelihood

4.1 Least favourable direction and target likelihood

Suppose that the parameterisation (ψ, χ) is not necessarily orthogonal. Let us consider an interest respecting reparameterisation leaving, for simplicity, the parameter ψ unchanged

$$(\psi, \chi) \longrightarrow (\psi, \lambda(\psi, \chi)) . \tag{6}$$

As a target likelihood for ψ, we take the likelihood for ψ with a known value of the nuisance parameter, in a suitably chosen reparameterisation of the form (6). In particular, we choose a parameterisation that gives the minimum information for ψ at each point (ψ, λ). Denoting by

$$l^{\lambda}(\psi) = l(\psi, \chi(\psi, \lambda))$$

the loglikelihood for ψ with λ fixed, we have

$$l^{\lambda}_{\psi} = l_{\psi} + \chi_{/\psi} l_{\chi} \tag{7}$$

$$l^{\lambda}_{\psi\psi} = l_{\psi\psi} + 2\chi_{/\psi} l_{\psi\chi} + (\chi_{/\psi})^2 l_{\chi\chi} + \chi_{/\psi\psi} l_{\chi} \tag{8}$$

$$i^{\lambda}_{\psi\psi} = i_{\psi\psi} + 2\chi_{/\psi} i_{\psi\chi} + (\chi_{/\psi})^2 i_{\chi\chi} , \tag{9}$$

where all quantities are evaluated at $(\psi, \chi(\psi, \lambda))$.

For each fixed value of ψ, $i^{\lambda}_{\psi\psi}$ depends only on the choice of the constant $\chi_{/\psi}$, and is minimum if

$$\chi_{/\psi} = -\beta , \tag{10}$$

where $\beta = i_{\psi\chi} i^{-1}_{\chi\chi}$. Condition (10) is the condition for orthogonality of ψ and λ (see for instance Barndorff-Nielsen & Cox, 1994, formula (2.74)). A loglikelihood for ψ, with λ fixed and with the reparameterisation (6) satisfying (10), will be denoted by $\bar{l}(\psi)$. It can be seen as the loglikelihood for ψ defined in the least favourable direction. From (7) and (10), the corresponding score function is $\bar{l}_\psi = l_\psi - \beta\, l_\chi$ and coincides with the efficient score function (Bickel et al., 1993, Section 2.4; see also Zhu & Reid, 1994). Moreover, (9) reduces to $\bar{\imath}_{\psi\psi} = i_{\psi\psi} - \beta\, i_{\psi\chi}$, which is the reciprocal of the Cramér–Rao lower bound for estimation of ψ, termed marginal Fisher information by Severini and Wong (1992) and partial information by Zhu & Reid (1994). Therefore, $\bar{l}(\psi)$ is an information preserving target loglikelihood. Note that when ψ and χ are orthogonal $\bar{l}(\psi)$ coincides with $l^\chi(\psi) = l(\psi, \chi)$.

The properties studied in this section hold for any choice of an orthogonal parameterisation and under any interest respecting reparameterisation. A key result is that $\bar{l}_\psi$ behaves as a tensor under such reparameterisations. Let $(\psi, \chi) \to (\phi, \xi)$, with $\phi = \phi(\psi)$ and $\xi = \xi(\psi, \chi)$, be a generic interest respecting reparameterisation. Let us use double overbars to denote likelihood quantities in this new parameterisation. From

$$\begin{aligned}
\bar{\bar{l}}_\phi &= l_\psi \psi_{/\phi} + l_\chi \chi_{/\phi} \\
\bar{\bar{l}}_\xi &= l_\chi \chi_{/\xi} \\
\bar{\bar{\imath}}_{\phi\xi} &= \left(i_{\psi\chi}\psi_{/\phi} + i_{\chi\chi}\chi_{/\phi}\right)\chi_{/\xi} \\
\bar{\bar{\imath}}_{\xi\xi} &= i_{\chi\chi}\left(\chi_{/\xi}\right)^2 ,
\end{aligned}$$

one concludes that the efficient score for ϕ in the new parameterisation, $\bar{\bar{l}}_\phi - \bar{\bar{\imath}}_{\phi\xi}\bar{\bar{\imath}}_{\xi\xi}^{-1}\bar{\bar{l}}_\xi$, is equal to $\bar{l}_\psi \psi_{/\phi}$.

4.2 Directed likelihoods: null and non-null distributions

Let $l_A(\psi)$ be an adjusted profile loglikelihood satisfying equation (1). Let us denote by $\bar{r}(\psi)$ the directed target likelihood, that is the directed likelihood computed from $\bar{l}(\psi)$.

The null distribution result of DiCiccio & Efron (1996) may be rewritten in terms of information preserving target likelihood. In particular,

$$E_\theta(r_A(\psi)) = \frac{1}{6\bar{\imath}_{\psi\psi}^{3/2}}\bar{t}_3(\psi) + O(n^{-3/2}) , \qquad (11)$$

$r_A(\psi)$ can be expanded in the form

$$\begin{aligned}
\mathrm{var}_\theta(r_A(\psi)) &= \mathrm{var}_\theta^\dagger(\bar{r}(\psi)) \\
&\quad +\bar{\imath}_{\psi\psi}^{-1}\bar{\imath}_{\lambda\lambda}^{-1}\left\{-\frac{3}{2}\bar{\nu}_{\psi\psi\lambda\lambda} - \bar{\nu}_{\psi\psi\lambda,\lambda} - \bar{\nu}_{\psi\lambda\lambda,\psi}\right\} \\
&\quad -\bar{\imath}_{\psi\psi}^{-2}\bar{\imath}_{\lambda\lambda}^{-1}\left\{\frac{1}{4}\bar{\nu}_{\psi\psi\lambda}^2 + \frac{1}{2}\bar{\nu}_{\psi\lambda\lambda}\bar{\nu}_{\psi\psi\psi} + \bar{\nu}_{\psi\psi\lambda}\bar{\nu}_{\psi\psi,\lambda} + \bar{\nu}_{\psi\lambda\lambda}\bar{\nu}_{\psi\psi,\psi}\right\} \\
&\quad -\bar{\imath}_{\psi\psi}^{-1}\bar{\imath}_{\lambda\lambda}^{-2}\left\{\frac{1}{4}\bar{\nu}_{\psi\lambda\lambda}^2 + \frac{1}{2}\bar{\nu}_{\psi\psi\lambda}\bar{\nu}_{\lambda\lambda\lambda} + \bar{\nu}_{\psi\lambda\lambda}\bar{\nu}_{\lambda\lambda,\psi} + \bar{\nu}_{\psi\psi\lambda}\bar{\nu}_{\lambda\lambda,\lambda}\right\} \\
&\quad +\bar{\imath}_{\psi\psi}^{-1}(\mu_\psi\mu_\psi - \mu_{\psi\psi}) + \bar{\imath}_{\psi\psi}^{-2}\mu_\psi(\bar{\nu}_{\psi\psi\psi} + 2\bar{\nu}_{\psi\psi,\psi}) \\
&\quad -\bar{\imath}_{\psi\psi}^{-1}\bar{\imath}_{\lambda\lambda}^{-1}\mu_\psi\bar{\nu}_{\psi\lambda\lambda} + 2\,\bar{\imath}_{\psi\psi}^{-1}\mu_{\psi/\psi} \\
&\quad +O(n^{-3/2}), \qquad (14)
\end{aligned}$$

where quantities such as $\bar{\nu}_{\psi\psi\lambda\lambda}$, $\bar{\nu}_{\psi\psi\lambda,\lambda}$ and so on, are the transformation of $\nu_{\psi\psi\chi\chi}$, $\nu_{\psi\psi\chi,\chi}$ and so on, under a reparameterisation (ψ, λ) satisfying (10).

The term of order $O(n^{-1})$ of the difference between $\mathrm{var}_\theta(r_A(\psi))$ and $\mathrm{var}_\theta^\dagger(\bar{r}(\psi))$ does not vanish in general. It does not vanish even when $l_A(\psi)$ has an information bias of order $O(n^{-1})$, as is true for the modified profile likelihood. In fact, for an $l_A(\psi)$ with information bias of order $O(n^{-1})$, the quantity $\mu_{\psi\psi}$ must be such that the leading term of (7) in DiCiccio et al. (1996) vanishes. In this case (14) simplifies as

$$\begin{aligned}
\mathrm{var}_\theta(r_A(\psi)) &= \mathrm{var}_\theta^\dagger(\bar{r}(\psi)) \\
&\quad -\bar{\imath}_{\psi\psi}^{-2}\bar{\imath}_{\lambda\lambda}^{-1}\left\{\frac{1}{4}\bar{\nu}_{\psi\psi\lambda}^2 + \bar{\nu}_{\psi\psi\lambda}\bar{\nu}_{\psi\psi,\lambda}\right\} + O(n^{-3/2}). \qquad (15)
\end{aligned}$$

For an orthogeodesic model, as defined in Barndorff-Nielsen & Blæsild (1993), we have that $\bar{\nu}_{\psi\psi\lambda} = 0$. As a consequence, $\mathrm{var}_\theta(r_A(\psi)) = \mathrm{var}_\theta(\bar{r}(\psi)) + O(n^{-3/2})$. Note that this is true for every interest respecting reparameterisation satisfying (10), and not only for the ortho-affine parameterisation (see Barndorff-Nielsen & Blæsild, 1993, p. 1023, property (iv)). Therefore, if the model is orthogeodesic and the information bias is of order $O(n^{-1})$, the Edgeworth expansion for the null distribution of $r_A(\psi)$ coincides with the Edgeworth expansion for the null distribution of $\bar{r}(\psi)$ up to terms of order $O(n^{-1})$ included. On the other hand, if the model is orthogeodesic, it can be shown that the information bias is of order $O(n^{-1})$ not only for the modified profile loglikelihood but also for the adjusted profile loglikelihoods of Cox & Reid (1987) and of Barndorff-Nielsen (1994).

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. (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *J. R. Statist. Soc.* B, **56**, 125-40.

BARNDORFF-NIELSEN, O. E. & BLÆSILD, P. (1993). Orthogeodesic models. *Ann. Statist.*, **21**, 1018-39.

BARNDORFF-NIELSEN, O. E. & COX, D. R. (1994). *Inference and Asymptotics*. London: Chapman and Hall.

BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. & WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.

COX, D. R. & REID, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc.* B, **49**, 1-39.

COX, D. R. & REID, N. (1989). On the stability of maximum-likelihood estimators of orthogonal parameters. *Can. J. Statist.*, **17**, 229-33.

DICICCIO, T. J. & EFRON, B. (1996). Bootstrap confidence intervals. *Statist. Sci.*, **11**, 189-228.

DICICCIO, T. J. & MARTIN, M. A. (1993). Simple modifications for signed roots of likelihood ratio statistics. *J. R. Statist. Soc.* B, **55**, 305-16.

DICICCIO, T. J. & STERN, S. E. (1994a). Constructing approximately standard normal pivots from signed roots of adjusted likelihood ratio statistics. *Scand. J. Statist.*, **21**, 447-60.

DICICCIO, T. J. & STERN, S. E. (1994b). Frequentist and Bayesian Bartlett corrections of test statistics based on adjusted profile likelihoods. *J. R. Statist. Soc.* B, **56**, 397-408.

DICICCIO, T. J., MARTIN, M. A., STERN, S. E. & YOUNG, G. A. (1996). Information bias and adjusted profile likelihoods. *J. R. Statist. Soc.* B, **58**, 189-203.

Lawley, D. N. (1956). A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika*, **43**, 295-303.

McCullagh, P. & Tibshirani, R. (1990). A simple method for the adjustement of profile likelihoods. *J. R. Statist. Soc.* B, **52**, 325-44.

Mukerjee, R. (1992). Comparison between the conditional likelihood ratio and the usual likelihood ratio tests. *J. R. Statist. Soc.* B, **54**, 189-94.

Sartori, N., Bellio, R., Salvan, A. & Pace, L. (1999). The directed modified profile likelihood in models with many nuisance parameters. *Biometrika*, **86**, 735-42.

Severini, T. (1998). Likelihood functions for inference in the presence of a nuisance parameter. *Biometrika*, **85**, 507-22.

Severini, T. & Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.*, **20**, 1768-802.

Stern, S. E. (1997). A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *J. R. Statist. Soc.* B, **59**, 653-65.

Zhu, Y. & Reid, N. (1994). Information, ancillarity, and sufficiency in the presence of nuisance parameters. *Can. J. Statist.*, **22**, 111-123.

Address for correspondence:

Alessandra Salvan
Department of Statistics
University of Padova
Via S. Francesco, 33
I-35121 Padova
Italy
salvan@stat.unipd.it