Adjustments of the profile likelihood from a new perspective

L. Pace, A. Salvan

2003.19

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

Ottobre 2003

Adjustments of the profile likelihood from a new perspective

Luigi Pace

Department of Statistics, University of Udine

Alessandra Salvan

Department of Statistics, University of Padova

October 20, 2003

Abstract

Various modifications of the profile likelihood have been proposed over the past twenty years. Their main theoretical basis is higher-order approximation of some target likelihood, defined by a suitable model reduction via conditioning or marginalisation, where the reduced model is indexed only by the parameter of interest. However, an exact reduced target likelihood exists only for special classes of models. In this paper, a general target likelihood is defined through model restriction along the least favourable curve in the parameter space. The profile likelihood can be seen as a purely estimative counterpart of this least favourable target likelihood. We will show that various modifications of the profile likelihood arise by refining the estimation process. In particular, bias reduction of the profile loglikelihood as an estimate of the expectation of the least favourable target loglikelihood gives adjustments that coincide to second order and agree with the available adjustments.

1 Introduction

Consider inference about a q-dimensional parameter of interest ψ in the presence of a nuisance parameter λ. Data y are the observation of an n-dimensional random variable Y with independent components and joint density $p_Y(y;\theta)$, where $\theta = (\psi, \lambda) \in \Theta \subseteq \mathbb{R}^d$. Standard first-order methods for inference about ψ are based on the profile likelihood. They can be seriously inaccurate when the dimension of λ is substantial relative to n. Various modifications of the profile likelihood have been proposed starting from Barndorff-Nielsen (1980, 1983), see Barndorff-Nielsen & Cox (1994, Chapter 8) and Severini (2000, Chapter 9) for accounts. All the available adjustments to the profile likelihood are equivalent to second order and share the common feature of reducing the score bias to $O(n^{-1})$ (DiCiccio *et al*, 1996).

Reduction of the score bias is the key basic motivation for adjusting the profile likelihood in McCullagh and Tibshirani (1990) and in Stern (1997). The other proposals, from Barndorff-Nielsen (1980, 1983) up to Fraser (2003), stem from a quite different approach. They aim to approximate some target likelihood, defined by a suitable model reduction via conditioning or marginalisation. Correction of the score bias is just a by-product of accurate approximation of a suitable genuine target likelihood.

Complementary to model reduction, there is another perspective for obtaining likelihood inference about an interest parameter. A target likelihood may also be defined through model restriction. This amounts to calculating the original likelihood along a curve $(\psi, \lambda(\psi))$ in the parameter space. One must ensure that no unrealistic information regarding ψ is introduced by this restriction. This leads to use of a least favourable curve (Stein, 1956), along which the Fisher information for ψ is equal to the adjusted (or partial) information for ψ in the original model. So far, model restriction along a least favourable curve seems to have been exploited mainly in the context of semiparametric and nonparametric models, leading to a generalization of the profile likelihood, as in Severini & Wong (1992) and in Murphy & van der Vaart (2000). But indeed the idea can be seen as lying at the basis of profile likelihood itself, see Severini (2000, Section 4.8). Here we will refer to the likelihood for ψ along the least favourable curve in the parameter space as the least favourable target likelihood.

However, the least favourable curve depends on the true parameter value, so that, unlike what happens with target likelihoods defined through model reduction, the least favourable target likelihood is not directly available for inference about ψ. Hence, a further estimation step is required. The profile likelihood can be seen as a purely estimative counterpart of the least favourable target likelihood (see Severini, 2000, Section 4.6). We will show that various modifications of the profile likelihood arise by refining the estimation process. This new rationale for modifications of the profile likelihood complements the results in Severini (1998a). Moreover, it suggests a fairly natural simulation approach for adjusting the profile likelihood.

A brief review of profile likelihood and its modifications is given in Section 2. Properties of the least favourable target likelihood are discussed in Section 3 together with its estimation. It turns out that, in the moderate-deviation neighbourhoods, the available adjusted profile loglikelihoods are equivalent to second order to a bias correction of the profile loglikelihood as an estimator of the expected least favourable target loglikelihood.

2 Profile likelihood and its modifications

Let us denote by $l(\theta) = l(\psi, \lambda) = l(\psi, \lambda; y) = \log p_Y(y; \theta)$ the full loglikelihood function and by $\hat{\theta} = (\hat{\psi}, \hat{\lambda})$ the maximum likelihood estimate of $\theta = (\psi, \lambda)$. Let $\hat{\lambda}_\psi$ be the constrained maximum likelihood estimate of λ for a given value of ψ and let $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$. The score vector $l_\theta = \frac{\partial}{\partial\theta} l(\theta)$ has blocks $l_\psi = l_\psi(\theta)$ and $l_\lambda = l_\lambda(\theta)$. Moreover, let $j_{\psi\psi} = j_{\psi\psi}(\theta)$, $j_{\psi\lambda} = j_{\psi\lambda}(\theta)$ and $j_{\lambda\lambda} = j_{\lambda\lambda}(\theta)$ be blocks of the observed information $j = j(\theta) = -\frac{\partial^2}{\partial\theta\partial\theta^\top} l(\theta)$. Similarly, we will denote by $i_{\psi\psi} = i_{\psi\psi}(\theta)$, $i_{\psi\lambda} = i_{\psi\lambda}(\theta)$ and $i_{\lambda\lambda} = i_{\lambda\lambda}(\theta)$ blocks of the expected information $i = i(\theta) = E_\theta(j(\theta))$.

Inference problems about ψ are invariant under interest respecting reparameterisations. An interest respecting reparameterisation is a reparameterisation $\omega = \omega(\theta) = (\varphi, \chi)$, where $\varphi = \varphi(\psi)$ and $\chi = \chi(\psi, \lambda)$ and, conversely, $\psi = \psi(\varphi)$ and $\lambda = \lambda(\varphi, \chi)$.

The profile loglikelihood is $l_P(\psi) = l(\hat{\theta}_\psi)$. Although $l_P(\psi)$ is not a genuine loglikelihood for ψ, it has many relevant properties as a surrogate loglikelihood. In particular, it is invariant under interest respecting reparameterisations, it is maximised by $\hat{\psi}$ and, under mild regularity conditions, the corresponding loglikelihood ratio statistic has asymptotic null distribution which is chi-squared with q degrees of freedom.

In small samples, the profile loglikelihood does not, however, take properly into account sampling variability of $\hat{\lambda}_\psi$. One effect of this is that the score computed from the profile likelihood typically has bias of order $O(1)$, that is

$$E_\theta\left(\frac{\partial l_P(\psi)}{\partial\psi}\right) = E_\theta\left(l_\psi(\psi, \hat{\lambda}_\psi)\right) = O(1)\,.$$

The same is true for the information bias:

$$E_\theta\left(\frac{\partial^2 l_P(\psi)}{\partial\psi\partial\psi^\top}\right) + E_\theta\left\{\left(\frac{\partial l_P(\psi)}{\partial\psi}\right)\left(\frac{\partial l_P(\psi)}{\partial\psi}\right)^\top\right\} = O(1)\,.$$

As a practical consequence, the usual chi-squared and normal approximations for the null distributions of the profile likelihood ratio statistic and of its signed version for a scalar ψ may be poor. Systematically misleading inferences are to be feared in particular when the dimension of λ is large relative to n.

When a suitable reduced marginal or conditional model exists whose densities depend only on ψ, inference about ψ may be based on the corresponding loglikelihood (Kalbfleisch & Sprott, 1970, 1973). A single approximating formula for these target loglikelihoods was obtained by Barndorff-Nielsen (1980, 1983). This is the modified profile loglikelihood, $l_M(\psi)$. Assume that the minimal sufficient statistic for the model is a one-to-one function of $(\hat{\psi}, \hat{\lambda}, a)$, where a is an ancillary statistic, either exactly or approximately, so that $l(\psi, \lambda; y) = l(\psi, \lambda; \hat{\psi}, \hat{\lambda}, a)$. Then,

$$l_M(\psi) = l_P(\psi) - \frac{1}{2} \log |\, j_{\lambda\lambda}(\hat{\theta}_\psi) \,| - \log \left| \frac{\partial \hat{\lambda}_\psi}{\partial \hat{\lambda}} \right| ,$$

where

$$\left| \frac{\partial \hat{\lambda}_\psi}{\partial \hat{\lambda}} \right| = \frac{|\, l_{\lambda;\hat{\lambda}}(\hat{\theta}_\psi) \,|}{|\, j_{\lambda\lambda}(\hat{\theta}_\psi) \,|} ,$$

involving the sample space derivatives $l_{\lambda;\hat{\lambda}}(\psi, \lambda) = \partial^2 l(\psi, \lambda; \hat{\psi}, \hat{\lambda}, a)/(\partial\lambda\, \partial\hat{\lambda}^\top)$. The modified profile likelihood has score bias of order $O(n^{-1})$. Its information bias is of order $O(n^{-1})$ as well (DiCiccio *et al.*, 1996). Calculation of sample space derivatives is straightforward only in special classes of models, notably exponential and group families.

When ψ and λ are orthogonal, i.e. $i_{\psi\lambda} = 0$, we have $\log \left| \partial\hat{\lambda}_\psi / \partial\hat{\lambda} \right| = O_p(n^{-1})$ in the moderate-deviation neighbourhoods, that is for $\psi - \hat{\psi} = O_p(n^{-1/2})$. Hence,

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

is an approximation of $l_M(\psi)$ with error of order $O_p(n^{-1})$ in the moderate-deviation neighbourhoods. The pseudologlikelihood $l_A(\psi)$ was proposed by Cox & Reid (1987). A major drawback of $l_A(\psi)$ is its lack of invariance under interest respecting reparameterisations.

Approximate calculation of sample space derivatives has been developed in Skovgaard (1996), Severini (1998b), Fraser, Reid & Wu (1999), Skovgaard (2001); see Severini (2000, Section 9.5) for a review. In particular, the approximation of $l_M(\psi)$ developed in Severini (1998b) is

$$\bar{l}_M(\psi) = l_P(\psi) + \frac{1}{2} \log |j_{\lambda\lambda}(\hat{\theta}_\psi)| - \log |\nu_{\lambda,\lambda}(\hat{\theta}_\psi, \hat{\theta}; \hat{\theta})| , \qquad (1)$$

where

$$\nu_{\lambda,\lambda}(\theta_1, \theta_2; \theta_0) = E_{\theta_0}(l_\lambda(\theta_1) l_\lambda(\theta_2)^\top) .$$

Analytical calculation of the adjustment term $\nu_{\lambda,\lambda}(\hat{\theta}_\psi, \hat{\theta}; \hat{\theta})$ is fairly simple in a number of important models (see e.g. Bellio and Brazzale, 2003). Monte Carlo calculation of the adjustment term is easy in broad generality.

Simulation results (DiCiccio & Martin, 1993; DiCiccio & Stern, 1994; 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 seems to be true in general and not only when a marginal or conditional target likelihood exists. Severini (1998a) provides a first theoretical justification of these empirical findings. He shows that, according to various asymptotic criteria, the modified profile likelihood is closer than the profile likelihood to a genuine likelihood. For a scalar parameter of interest, a different but related justification is provided in Sartori *et al.* (2003). It is shown there that the null and non-null distributions of the directed likelihood calculated from an adjusted profile likelihood having score bias of order $O(n^{-1})$ are closer than the corresponding distributions of the directed profile likelihood to the distributions of a directed likelihood calculated from a genuine likelihood. A theoretical justification for the remarkable accuracy of inference based on the modified profile likelihood in the presence of many nuisance parameters is provided by Sartori (2003) using a two-index asymptotics setting.

3 The least favourable target likelihood and bias correction of the profile loglikelihood

3.1 Partial information

If λ were known to be equal to λ_0, likelihood inference about ψ would be based on $l(\psi, \lambda_0)$. Estimation of λ_0 with $\hat{\lambda}$ and use of $l(\psi, \hat{\lambda})$ as a pseudologlikelihood for

ψ (Gong and Samaniego, 1981) overrates information about ψ unless ψ and λ are orthogonal. Indeed, the partial expected information on ψ is

$$i_{\psi\psi\cdot\lambda}(\theta) = i_{\psi\psi}(\theta) - i_{\psi\lambda}(\theta) i_{\lambda\lambda}(\theta)^{-1} i_{\lambda\psi}(\theta) \, .$$

Note that $i_{\psi\psi\cdot\lambda}(\theta)$ is the inverse of the Cramér-Rao lower bound for estimation of ψ when λ is unknown. Hence $l(\psi, \hat{\lambda})$ usually overrates information because the expected information on ψ computed from $l(\psi, \hat{\lambda})$ is first-order equivalent to $i_{\psi\psi}(\theta)$.

On the other hand, the profile loglikelihood preserves information about ψ. The observed information from $l_P(\psi)$ is

$$-\frac{\partial^2}{\partial\psi\partial\psi^\top} l_P(\psi) = j_{\psi\psi}(\hat{\theta}_\psi) - j_{\psi\lambda}(\hat{\theta}_\psi) j_{\lambda\lambda}(\hat{\theta}_\psi)^{-1} j_{\lambda\psi}(\hat{\theta}_\psi) \, ,$$

which, at the true ψ, is first-order equivalent to $i_{\psi\psi\cdot\lambda}(\theta)$.

3.2 The least favourable target likelihood

Let $\theta_0 = (\psi_0, \lambda_0)$ denote the true parameter value and let $E_0(\cdot)$ and $V_0(\cdot)$ denote, respectively, expectation and variance under θ_0. Under regularity conditions, $\hat{\lambda}_\psi$ is a consistent estimator of λ_ψ, the maximiser of $E_0(l(\psi, \lambda))$ with respect to λ for fixed ψ (Huber, 1967). The least favourable target loglikelihood is defined here as

$$l_T(\psi) = l(\theta_\psi) \, ,$$

where $\theta_\psi = (\psi, \lambda_\psi)$. Note that $l_T(\psi)$ is a genuine loglikelihood, but it is not available in practice, because λ_ψ depends on θ_0. When needed, we will use the more explicit notation $\lambda_\psi = \lambda(\psi; \theta_0) = \lambda(\psi; \psi_0, \lambda_0)$.

The curve θ_ψ is a least favourable curve in the parameter space, according to the definition of Stein (1956). Indeed, Fisher information about ψ calculated from $l(\psi, \lambda_\psi)$ evaluated at θ_0 coincides with $i_{\psi\psi\cdot\lambda}(\theta_0)$. See also Severini & Wong (1992, Section 4) and Severini (2000, Sections 4.6 and 4.8).

The curve θ_ψ is least favourable even according to the following global sense. Let $\tilde{\lambda}_\psi$ be a function of ψ such that $\tilde{\lambda}_{\psi_0} = \lambda_0$ and let $l_R(\psi) = l(\psi, \tilde{\lambda}_\psi)$ be a generic loglikelihood for ψ obtained through model restriction. Then

$$E_0\{l_R(\psi_0) - l_R(\psi)\} \geq E_0\{l_T(\psi_0) - l_T(\psi)\}\,.$$

Hence, for any given $\psi \neq \psi_0$, the curve θ_ψ minimises Kullback-Leibler divergence between $p_Y(y;\theta_0)$ and $p_Y(y;\psi,\tilde{\lambda}_\psi)$ among all possible curves $(\psi, \tilde{\lambda}_\psi)$ with $\tilde{\lambda}_{\psi_0} = \lambda_0$.

Finally, λ_ψ is related to a locally orthogonal reparameterisation. Any $\tilde{\lambda}_\psi$ can be thought of as obtained from a reparameterisation $(\psi, \chi(\psi,\lambda))$, by equating χ to $\chi_0 = \chi(\psi_0, \lambda_0)$. Fisher information about ψ calculated from $l(\psi, \tilde{\lambda}_\psi)$ and evaluated at θ_0 coincides with $i_{\psi\psi\cdot\lambda}(\theta_0)$ if and only if the reparameterisation $(\psi, \chi(\psi, \lambda))$ is locally orthogonal at θ_0.

3.3 Bias correction of the profile loglikelihood

A desirable property for a pseudologlikelihood $l_{PS}(\psi)$ is that it be an unbiased estimate of $E_0(l_T(\psi))$. Indeed, suppose that for every $\theta_0 \in \Theta$ we have $E_0(l_{PS}(\psi)) = E_0(l_T(\psi))$. Then, as a first consequence, $l_{PS}(\psi)$ has some properties of a genuine likelihood. In particular, it satisfies Wald inequality $E_0(l_{PS}(\psi_0)) > E_0(l_{PS}(\psi))$, for $\psi \neq \psi_0$, and $(\partial/\partial\psi) l_{PS}(\psi)$ is an unbiased estimating function for ψ. Moreover, at the true ψ, the expected curvature of such a pseudologlikelihood gives the right information. In other words, minus the expected Hessian matrix from $l_{PS}(\psi)$ at $\psi = \psi_0$ coincides with $i_{\psi\psi\cdot\lambda}(\theta_0)$, the partial expected information on ψ.

It is clear that $l_P(\psi)$ is the estimative, or plug-in, counterpart of $l_T(\psi)$ and has a bias of order $O(1)$ as an estimate of $E_0(l_T(\psi))$. Therefore we aim at an adjusted estimative counterpart of $l_T(\psi)$, $l_{AE}(\psi)$, which has to be of the form

$$l_{AE}(\psi) = l_P(\psi) - a(\psi)\,, \tag{2}$$

where the adjustment $a(\psi)$ is an estimate of the bias term

$$b(\psi;\theta_0) = E_0\left(l_P(\psi) - l_T(\psi)\right) .$$

Various asymptotically equivalent approximations may be obtained for the bias term and its estimate, leading to asymptotically equivalent versions of $l_{AE}(\psi)$. In particular, we will show that three asymptotically equivalent versions of $l_{AE}(\psi)$ are

$$l^{I}_{AE}(\psi) = l_P(\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\hat\theta_\psi)| - \frac{1}{2}\log|V_{\hat\theta}(\hat\lambda_\psi)| , \tag{3}$$

$$l^{II}_{AE}(\psi) = l_P(\psi) + \frac{1}{2}\log|j_{\lambda\lambda}(\hat\theta_\psi)| - \frac{1}{2}\log|\nu_{\lambda,\lambda}(\hat\theta_\psi,\hat\theta_\psi;\hat\theta)| , \tag{4}$$

$$l^{III}_{AE}(\psi) = l_P(\psi) + \frac{1}{2}\log|j_{\lambda\lambda}(\hat\theta_\psi)| - \log|\nu_{\lambda,\lambda}(\hat\theta_\psi,\hat\theta;\hat\theta)| . \tag{5}$$

Many available adjustments to the profile loglikelihood may be seen as connected to the above versions of $l_{AE}(\psi)$.

When ψ and λ are orthogonal, $\hat\lambda_\psi = \hat\lambda + O(n^{-1})$ so that $l^{I}_{AE}(\psi) = l_A(\psi) + O_p(n^{-1})$. Thus, the term $-\frac{1}{2}\log|V_{\hat\theta}(\hat\lambda_\psi)|$ accounts for nonorthogonality of ψ and λ. However, like $l_A(\psi)$, $l^{I}_{AE}(\psi)$ is clearly not invariant under interest respecting reparameterisations.

Versions (4) and (5) are invariant under interest respecting reparameterisations. In a multiparameter exponential family with loglikelihood $l(\psi,\lambda) = \psi\cdot t + \lambda\cdot u - nK(\psi,\lambda)$, we have $j_{\lambda\lambda}(\hat\theta_\psi) = nK_{\lambda\lambda}(\hat\theta_\psi)$ and $\nu_{\lambda,\lambda}(\hat\theta_\psi,\hat\theta_\psi;\hat\theta) = \nu_{\lambda,\lambda}(\hat\theta,\hat\theta_\psi;\hat\theta) = nK_{\lambda\lambda}(\hat\theta)$, so that $l^{II}_{AE}(\psi) = l^{III}_{AE}(\psi) = l_M(\psi)$.

An empirical analogue of $\nu_{\lambda,\lambda}(\hat\theta_\psi,\hat\theta_\psi;\hat\theta)$ was used by DiCiccio & Stern (1993, Stanford Technical Report) for a modification of $l_P(\psi)$ closely related to $l^{II}_{AE}(\psi)$.

Finally, notice that $l^{III}_{AE}(\psi)$ coincides with Severini (1998b) approximation (1) of $l_M(\psi)$. This is the main result of the paper, showing that bias reduction of the profile loglikelihood as an estimate of the expected target loglikelihood is a useful perspective for adjustments of the profile loglikelihood.

We sketch below the derivation of (3)–(5). For simplicity, we assume in the following that both ψ and λ are scalars. Moreover, we let $i_{\lambda\lambda}(\theta_\psi;\theta_0) = E_0(j_{\lambda\lambda}(\theta_\psi))$.

Various asymptotically equivalent approximations may be obtained for the bias term. Standard asymptotic expansions (see Appendix A) give

$$\begin{aligned} b(\psi;\theta_0) &= b^I(\psi;\theta_0)+O(n^{-1}) \\ &= b^{II}(\psi;\theta_0)+O(n^{-1}) \\ &= b^{III}(\psi;\theta_0)+O(n^{-1})\,, \end{aligned}$$

where

$$b^I(\psi;\theta_0) = \frac{1}{2} i_{\lambda\lambda}(\theta_\psi;\theta_0) V_0(\hat{\lambda}_\psi)\,, \tag{6}$$

$$b^{II}(\psi;\theta_0) = \frac{1}{2} i_{\lambda\lambda}(\theta_\psi;\theta_0)^{-1} \nu_{\lambda,\lambda}(\theta_\psi,\theta_\psi;\theta_0)\,, \tag{7}$$

$$b^{III}(\psi;\theta_0) = \frac{1}{2} \nu_{\lambda,\lambda}(\theta_\psi,\theta_0;\theta_0)^2 i_{\lambda\lambda}(\theta_0)^{-1} i_{\lambda\lambda}(\theta_\psi;\theta_0)^{-1}\,. \tag{8}$$

While it is clear that $b^I(\psi;\theta_0)$ is not invariant under interest respecting reparameterisations, $b^{II}(\psi;\theta_0)$ and $b^{III}(\psi;\theta_0)$ are invariant under such reparameterisations.

Consistent estimates of $b^I(\psi;\theta_0)$, $b^{II}(\psi;\theta_0)$ and $b^{III}(\psi;\theta_0)$ are obtained by replacing $i_{\lambda\lambda}(\theta_\psi;\theta_0)$ with $j_{\lambda\lambda}(\theta_\psi)$, θ_ψ with $\hat{\theta}_\psi$ and θ_0 with $\hat{\theta}$ in (6)–(8). Therefore, we have the estimates of the bias term

$$a^I(\psi) = \frac{1}{2} j_{\lambda\lambda}(\hat{\theta}_\psi) V_{\hat{\theta}}(\hat{\lambda}_\psi)\,, \tag{9}$$

$$\begin{aligned} a^{II}(\psi) &= \frac{1}{2} j_{\lambda\lambda}(\hat{\theta}_\psi)^{-1} \nu_{\lambda,\lambda}(\hat{\theta}_\psi,\hat{\theta}_\psi;\hat{\theta})\,, \\ a^{III}(\psi) &= \frac{1}{2} \nu_{\lambda,\lambda}(\hat{\theta}_\psi,\hat{\theta};\hat{\theta})^2 j_{\lambda\lambda}(\hat{\theta})^{-1} j_{\lambda\lambda}(\hat{\theta}_\psi)^{-1}\,. \end{aligned}$$

The adjusted estimative loglikelihoods (3)–(5) are obtained by considering the following local versions of $a^I(\psi)$, $a^{II}(\psi)$ and $a^{III}(\psi)$, with error of order $O(n^{-1})$ for $\psi-\hat{\psi}=O(n^{-1/2})$,

$$\begin{aligned} \tilde{a}^I(\psi) &= \frac{1}{2}\log j_{\lambda\lambda}(\hat{\theta}_\psi) + \frac{1}{2}\log V_{\hat{\theta}}(\hat{\lambda}_\psi)\,, \\ \tilde{a}^{II}(\psi) &= -\frac{1}{2}\log j_{\lambda\lambda}(\hat{\theta}_\psi) + \frac{1}{2}\log \nu_{\lambda,\lambda}(\hat{\theta}_\psi,\hat{\theta}_\psi;\hat{\theta})\,, \\ \tilde{a}^{III}_{LAE}(\psi) &= -\frac{1}{2}\log j_{\lambda\lambda}(\hat{\theta}_\psi) + \log \nu_{\lambda,\lambda}(\hat{\theta}_\psi,\hat{\theta};\hat{\theta})\,. \end{aligned}$$

The local version $\tilde{a}^I(\psi)$ is obtained from $a^I(\psi)$ using the expansion

$$\begin{aligned}\frac{1}{2}j_{\lambda\lambda}(\hat{\theta}_\psi)V_{\hat{\theta}}(\hat{\lambda}_\psi) &= \frac{1}{2}\left(1 + j_{\lambda\lambda}(\hat{\theta}_\psi)V_{\hat{\theta}}(\hat{\lambda}_\psi) - 1\right)\\ &= \frac{1}{2}\left\{1 + \log\left(1 + j_{\lambda\lambda}(\hat{\theta}_\psi)V_{\hat{\theta}}(\hat{\lambda}_\psi) - 1\right)\right\} + O(n^{-1})\\ &= \frac{1}{2}\log j_{\lambda\lambda}(\hat{\theta}_\psi) + \frac{1}{2}\log V_{\hat{\theta}}(\hat{\lambda}_\psi) + \text{constant} + O(n^{-1}).\end{aligned}$$

The local versions $\tilde{a}^{II}(\psi)$ and $\tilde{a}^{III}(\psi)$ are obtained in a similar way.

As a final remark, thinking of adjusted profile loglikelihood as a bias correction of the profile loglikelihood suggests a fairly natural simulation approach for adjusting the profile likelihood.

Parametric bootstrap samples $y^*(r)$, generated from the model with density $p_Y(y;\hat{\theta})$, are used to calculate constrained maximum likelihood estimates $\hat{\lambda}^*_\psi(r)$, $r = 1, \dots, R$. The bootstrap adjusted estimative loglikelihood is defined as

$$l_{BAE}(\psi) = \frac{1}{R}\sum_{r=1}^{R} l(\psi, \hat{\lambda}^*_\psi(r); y).$$

Note that $l_{BAE}(\psi)$ is invariant under interest respecting reparameterisations. Moreover, involving only constrained maximisation, it does not require an explicit nuisance parameterisation. It is straightforward to check through expansions that $l_{BAE}(\psi)$ is asymptotically of the form (2) with $a(\psi)$ given by (9). Indeed, in a moderate-deviation region,

$$l(\psi, \hat{\lambda}^*_\psi(r); y) = l(\psi, \hat{\lambda}_\psi; y) + (\hat{\lambda}^*_\psi(r) - \hat{\lambda}_\psi)l_\lambda(\psi, \hat{\lambda}_\psi; y) + \frac{1}{2}(\hat{\lambda}^*_\psi(r) - \hat{\lambda}_\psi)^2 l_{\lambda\lambda}(\psi, \hat{\lambda}_\psi; y) + O_p(n^{-1}),$$

where $l_\lambda(\psi, \hat{\lambda}_\psi; y) = 0$, so that

$$l_{BAE}(\psi) = l_P(\psi) - \frac{1}{2}j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)\frac{1}{R}\sum_{r=1}^{R}(\hat{\lambda}^*_\psi(r) - \hat{\lambda}_\psi)^2 + O_p(n^{-1}),$$

where $\frac{1}{R}\sum_{r=1}^{R}(\hat{\lambda}^*_\psi(r) - \hat{\lambda}_\psi)^2$ is the empirical analogue of $V_{\hat{\theta}}(\hat{\lambda}_\psi)$.

Appendix A

We show first that

$$b(\psi;\theta_0) = E_0\left(l_{_P}(\psi) - l_{_T}(\psi)\right) = b^I(\psi;\theta_0) + O(n^{-1}),$$

with $b^I(\psi;\theta_0)$ given by (6).

Let $l_{\lambda\lambda}(\theta) = (\partial^2/\partial\lambda^2)l(\theta)$ and $l_{\lambda\lambda\lambda}(\theta) = (\partial^3/\partial\lambda^3)l(\theta)$. Consider the expansion

$$l_{_T}(\psi) = l_{_P}(\psi) + \frac{1}{2}(\hat{\lambda}_\psi - \lambda_\psi)^2 l_{\lambda\lambda}(\hat{\theta}_\psi) - \frac{1}{6}(\hat{\lambda}_\psi - \lambda_\psi)^3 l_{\lambda\lambda\lambda}(\hat{\theta}_\psi) + O_p(n^{-1}),$$

which gives, after further expansions,

$$\begin{aligned} l_{_P}(\psi) &= l_{_T}(\psi) + \frac{1}{2}(\hat{\lambda}_\psi - \lambda_\psi)^2 i_{\lambda\lambda}(\theta_\psi;\theta_0) \\ &\quad -\frac{1}{2}(\hat{\lambda}_\psi - \lambda_\psi)^2 H_{\lambda\lambda}(\theta_\psi;\theta_0) - \frac{1}{3}(\hat{\lambda}_\psi - \lambda_\psi)^3 \nu_{\lambda\lambda\lambda}(\theta_\psi;\theta_0) + O_p(n^{-1}), \quad (10) \end{aligned}$$

where $H_{\lambda\lambda}(\theta_\psi;\theta_0) = -j_{\lambda\lambda}(\theta_\psi) + i_{\lambda\lambda}(\theta_\psi;\theta_0)$ and $\nu_{\lambda\lambda\lambda}(\theta_\psi;\theta_0) = E_0(l_{\lambda\lambda\lambda}(\theta_\psi))$. Hence,

$$E_0(l_{_P}(\psi)) = E_0(l_{_T}(\psi)) + \frac{1}{2} i_{\lambda\lambda}(\theta_\psi;\theta_0) V_0(\hat{\lambda}_\psi) + O(n^{-1}). \quad (11)$$

We now show that

$$b(\psi;\theta_0) = E_0\left(l_{_P}(\psi) - l_{_T}(\psi)\right) = b^{II}(\psi;\theta_0) + O(n^{-1}),$$

with $b^{II}(\psi;\theta_0)$ given by (7).

From $l_\lambda(\hat{\theta}_\psi) = 0$, in a standard way, the following expansion for $\hat{\lambda}_\psi - \lambda_\psi$ is obtained:

$$\begin{aligned} \hat{\lambda}_\psi - \lambda_\psi &= i_{\lambda\lambda}(\theta_\psi;\theta_0)^{-1} l_\lambda(\theta_\psi) + i_{\lambda\lambda}(\theta_\psi;\theta_0)^{-2} H_{\lambda\lambda}(\theta_\psi;\theta_0) l_\lambda(\theta_\psi) \\ &\quad + \frac{1}{2} i_{\lambda\lambda}(\theta_\psi;\theta_0)^{-3} \nu_{\lambda\lambda\lambda}(\theta_\psi;\theta_0) l_\lambda(\theta_\psi)^2 + O_p(n^{-3/2}). \end{aligned}$$

Notice that $E_0(l_\lambda(\theta_\psi)) = 0$ due to the maximising property of λ_ψ. The above expansion for $\hat{\lambda}_\psi - \lambda_\psi$ gives $V_0(\hat{\lambda}_\psi) = i_{\lambda\lambda}(\theta_\psi;\theta_0)^{-2} \nu_{\lambda,\lambda}(\theta_\psi,\theta_\psi;\theta_0) + O(n^{-2})$.

Finally, we show that

$$b(\psi;\theta_0) = E_0\left(l_{_P}(\psi) - l_{_T}(\psi)\right) = b^{III}(\psi;\theta_0) + O(n^{-1}),$$

with $b^{III}(\psi;\theta_0)$ given by (8). In particular, we will show that $b^{III}(\psi;\theta_0) = b^{II}(\psi;\theta_0) + O(n^{-1})$ for $\psi - \psi_0 = O(n^{-1/2})$.

For $\psi - \psi_0 = O(n^{-1/2})$, consider the expansion

$$l_\lambda(\theta_\psi) = l_\lambda(\theta_0) + l_{\lambda\psi}(\theta_0)(\psi - \psi_0) + l_{\lambda\lambda}(\theta_0)(\lambda_\psi - \lambda_0) + O_p(1)\,.$$

Hence,

$$\nu_{\lambda,\lambda}(\theta_\psi, \theta_0; \theta_0) = i_{\lambda\lambda}(\theta_0) + \nu_{\lambda,\lambda\psi}(\theta_0)(\psi - \psi_0) + \nu_{\lambda,\lambda\lambda}(\theta_0)(\lambda_\psi - \lambda_0) + O(1)\,.$$

Similarly,

$$\nu_{\lambda,\lambda}(\theta_\psi, \theta_\psi; \theta_0) = i_{\lambda\lambda}(\theta_0) + 2\nu_{\lambda,\lambda\psi}(\theta_0)(\psi - \psi_0) + 2\nu_{\lambda,\lambda\lambda}(\theta_0)(\lambda_\psi - \lambda_0) + O(1)\,.$$

As a consequence,

$$i_{\lambda\lambda}(\theta_0)^{-2}\nu_{\lambda,\lambda}(\theta_\psi, \theta_0; \theta_0)^2 = i_{\lambda\lambda}(\theta_0)^{-1}\nu_{\lambda,\lambda}(\theta_\psi, \theta_\psi; \theta_0) + O(n^{-1})\,.$$

References

BARNDORFF-NIELSEN, O. E. (1980). Conditionality resolutions. *Biometrika*, **67**, 293-310.

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

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

BELLIO, R. & BRAZZALE, A.R. (2003). Higher-order asymptotics unleashed: Software design for nonlinear heteroscedastic models. *J. Comput. Graph. Statist.*, **12**, 682-697.

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

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

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

Fraser, D. A. S. (2003). Likelihood for component parameters. *Biometrika*, **?**, ?.

Fraser, D. A. S., Reid, N. & Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika*, **86**, 249-64.

Gong, G. & Samaniego, F.J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Ann. Statist.*, **89**, 861-9.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.*, **1**, 221-34, Univ. California Press, Berkeley.

Kalbfleisch, J. D. & Sprott, D. A. (1970). Applications of likelihood methods to models involving a large number of parameters (with discussion). *J. Roy. Statist. Soc.* B, **32**, 175-208.

Kalbfleisch, J. D. & Sprott, D. A. (1973). Marginal and conditional likelihoods. *Sankhyā* A, **35**, 311-28.

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

Murphy, S. A. & van der Vaart, A. W. (2000). On profile likelihood. *J. Amer. Statist. Assoc.*, **95**, 449-65

Sartori, N. (2003). Modified profile likelihoods in models with stratum nuisance parameters. *Biometrika*, **90**, 533-49.

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

Sartori, N., Salvan, A. & Pace, L. (2003). A note on directed adjusted profile likelihoods. *J. Stat. Plann. Inf.*, **110**, 1-9.

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

Severini, T. (1998b). An approximation to the modified profile likelihood function. *Biometrika*, **85**, 403-11.

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

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

SKOVGAARD, I. M. (1996). An explicit large-deviation approximation to one-parameter tests. *Bernoulli*, **2**, 145-65.

SKOVGAARD, I. M. (2001). Likelihood asymptotics. *Scand. J. Statist.*, **28**, 3-32.

STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Probab.*, **1**, 187-95, Univ. California Press, Berkeley.

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.