An adjustment of the modified profile likelihood from matching priors

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Walter Racugno
Department of Mathematics
University of Cagliari
Italy

Abstract: Various modifications of the profile likelihood have been proposed in the literature. Despite modified profile likelihood methods have better properties than those based on the profile likelihood, the signed likelihood ratio statistic based on the modified profile likelihood has a standard normal distribution only to first order, and it can be inaccurate in particular in models with many nuisance parameters. In this paper we propose an adjustment of the profile likelihood from a new perspective. The idea is to resort to suitable default priors on the parameter of interest only to be used as non-negative weight functions in order to modify the modified profile likelihood. In particular, we focus on matching priors, i.e. priors on the parameter of interest only for which there is an agreement between frequentist and Bayesian inference, derived from modified profile likelihoods. The proposed modified profile likelihood has desiderable inferential properties: the corresponding signed likelihood ratio statistic is standard normal to second order and the correponding maximizer is a refinement of the maximum likelihood estimator, which improves its small sample properties. Examples illustrate the proposed modified profile likelihood and outline its improvement over its counterparts.

Keywords: Bayesian inference; Exponential family; Group model; Higher-order asymptotics; Modified profile likelihood; Nuisance parameter; Skew-normal distribution; Signed and modified signed likelihood ratio statistic.

Contents

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:
Laura Ventura
tel: +39 049 827 4177
ventura@stat.unipd.it
http://homes.stat.unipd.it/ventura/

An adjustment of the modified profile likelihood from matching priors

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Walter Racugno
Department of Mathematics
University of Cagliari
Italy

Abstract: Various modifications of the profile likelihood have been proposed in the literature. Despite modified profile likelihood methods have better properties than those based on the profile likelihood, the signed likelihood ratio statistic based on the modified profile likelihood has a standard normal distribution only to first order, and it can be inaccurate in particular in models with many nuisance parameters. In this paper we propose an adjustment of the profile likelihood from a new perspective. The idea is to resort to suitable default priors on the parameter of interest only to be used as non-negative weight functions in order to modify the modified profile likelihood. In particular, we focus on matching priors, i.e. priors on the parameter of interest only for which there is an agreement between frequentist and Bayesian inference, derived from modified profile likelihoods. The proposed modified profile likelihood has desiderable inferential properties: the corresponding signed likelihood ratio statistic is standard normal to second order and the correponding maximizer is a refinement of the maximum likelihood estimator, which improves its small sample properties. Examples illustrate the proposed modified profile likelihood and outline its improvement over its counterparts.

Keywords: Bayesian inference; Exponential family; Group model; Higher-order asymptotics; Modified profile likelihood; Nuisance parameter; Skew-normal distribution; Signed and modified signed likelihood ratio statistic.

1 Introduction

Let us consider a model with a scalar parameter of interest ψ, a d-dimensional nuisance parameter λ and likelihood function $L(\psi, \lambda) = L(\psi, \lambda; y)$, where $y = (y_1, \ldots, y_n)$ is a random sample of size n. Standard first-order methods for inference about ψ are based on the profile likelihood $L_p(\psi) = L(\psi, \hat{\lambda}_\psi)$, where $\hat{\lambda}_\psi$ is the maximum likelihood estimator of λ for fixed ψ, and can be seriously inaccurate, in particular when the dimension of λ is substantial relative to n. Starting from

Barndorff-Nielsen (1980, 1983), various modifications of the profile likelihood of the form

$$L_{mp}(\psi) = L_p(\psi)\, M(\psi) \tag{1}$$

have been proposed, for suitably defined correction terms $M(\psi)$; see Barndorff-Nielsen and Cox (1994, Chapter 8) and Severini (2000, Chapter 9) for detailed accounts. Reduction of the score bias is the key basic motivation for adjusting $L_p(\psi)$ in McCullagh and Tibshirani (1990) and in Stern (1997), while the other proposals, from Barndorff-Nielsen (1980, 1983) up to Fraser (2003) and Pace and Salvan (2006), aim to approximate some target likelihood. All the available adjustments to the profile likelihood are equivalent to second order, share the common feature of reducing the score bias to $O(n^{-1})$ (DiCiccio *et al.*, 1996), and in general provide accurate inferences on ψ. However, the signed likelihood ratio statistic based on $L_{mp}(\psi)$ is standard normal only to first order, and it can be inaccurate in particular in models with many nuisance parameters (Sartori *et al.*, 1999).

The aim of this paper is to discuss a modification of the profile likelihood from a new perspective, based on a non standard frequentist approach which makes use of Bayesian inferential procedures. More precisely, we use a suitable default prior on the parameter of interest only, which can be interpreted as a non-negative weight function on ψ, as a convenient device for adjusting the modified profile likelihood of Barndorff-Nielsen (1983). The possibility of adjusting a likelihood function using priors, even if quite differently motivated, is suggested also in Efron (1993), Liseo (1993), and Reid (1995). Here, we focus on the class of strong matching priors for ψ only derived from $L_{mp}(\psi)$ (Ventura *et al.*, 2009, Ventura and Racugno, 2010), i.e. priors for which there is an agreement between frequentist and Bayesian results and which validate the use of $L_{mp}(\psi)$ for Bayesian inference in the presence of nuisance parameters. We then suggest to modify $L_{mp}(\psi)$ with its corresponding matching prior, given by

$$\pi(\psi) \propto i_{\psi\psi.\lambda}(\psi, \hat{\lambda}_\psi)^{1/2} \ , \tag{2}$$

where $i_{\psi\psi.\lambda}(\psi,\lambda) = i_{\psi\psi}(\psi,\lambda) - i_{\psi\lambda}(\psi,\lambda) i_{\lambda\lambda}(\psi,\lambda)^{-1} i_{\lambda\psi}(\psi,\lambda)$ is the partial information, with $i_{\psi\psi}(\psi,\lambda)$, $i_{\psi\lambda}(\psi,\lambda)$, $i_{\lambda\lambda}(\psi,\lambda)$, and $i_{\lambda\psi}(\psi,\lambda)$ blocks of the expected Fisher information $i(\psi,\lambda)$ from $L(\psi,\lambda)$. The implied modified profile likelihood is thus defined as

$$L^*_{mp}(\psi) = L_{mp}(\psi)\ i_{\psi\psi.\lambda}(\psi, \hat{\lambda}_\psi)^{1/2} \ . \tag{3}$$

We will show that $L^*_{mp}(\psi)$ has better inferential properties than $L_{mp}(\psi)$. In particular, for tests or for confidence intervals, the signed likelihood ratio statistic based on (3) is standard normal to second order, giving quite accuarate inferences also for small sample sizes. Moreover, for point estimation, the maximizer of (3) is the solution of an estimating equation obtained from an higher-order pivot for the parameter of interest, i.e. it is a refinement of the maximum likelihood estimator $\hat{\psi}$, improving some small sample properties and keeping equivariance under reparameterisation (Pace and Salvan, 1999, Giummolé and Ventura, 2002).

The plan of the paper is as follows. Section 2 gives some background on matching priors derived from modified profile likelihoods and on signed likelihood ratio statistics based on $L_p(\psi)$ and $L_{mp}(\psi)$. In Section 3 it is shown that the signed likelihood ratio statistic based on (3) is standard normal to second order, and is equivalent to the modified signed likelihood ratio statistic of Barndorff-Nielsen and Chamberlin (1994). Moreover, the properties of the maximizer of (3) are highlighted. In Section 3 simulation results are presented to confirm that the proposed modified profile likelihood improves on its counterparts.

2 Background theory

Bayesian versus frequentist interface in statistical inference has been the subject of considerable recent interest. For instance, this has lead to the investigation of integrated likelihood functions for non-Bayesian inference (Liseo, 1993, Berger *et al.*, 1999, Severini, 2007, 2010, 2011), to the development of matching priors that ensure approximate frequentist validity of posterior credible regions (see e.g. Datta and Mukerjee, 2004), to the use Bayesian expansions for frequentist computations (see, for instance, Mukerjee and Reid, 2000), and to the use of pseudo-likelihood functions for Bayesian inference (see, among others, Severini, 1999, Lazar, 2003, Chang and Mukerjee, 2006, Lin, 2006, Greco *et al.*, 2008, Ventura *et al.*, 2009, 2010, Racugno *et al.*, 2010, Pauli *et al.*, 2011, and references therein).

In particular, the agreement between the frequentist and posterior coverage probabilities of credible regions, arising from matching priors on ψ only derived from $L_{mp}(\psi)$, provides a validation for these priors, and hence their study is of interest from the frequentist viewpoint as well. Let us consider the modified profile likelihood of Barndorff-Nielsen (1983), given by

$$L_{mp}(\psi) = L_p(\psi)\, C(\psi) \ , \tag{4}$$

where

$$C(\psi) = \frac{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2} |j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}}{|\ell_{\lambda;\hat{\lambda}}(\psi, \hat{\lambda}_\psi)|} \ , \tag{5}$$

$j_{\lambda\lambda}(\psi, \lambda)$ is the (λ, λ)-block of the observed Fisher information $j(\psi, \lambda)$, $\ell_{\lambda;\hat{\lambda}}(\psi, \lambda) = \partial \ell(\psi, \lambda)/\partial\lambda\partial\hat{\lambda}^\mathsf{T}$, with $\ell(\psi, \lambda) = \log L(\psi, \lambda)$, is a sample space derivative, and $(\hat{\psi}, \hat{\lambda})$ is the maximum likelihood estimator of (ψ, λ). The modified profile likelihood (4) depends only on the data and the parameter of interest ψ and thus it can be used also in the Bayesian framework as a genuine likelihood function to construct a posterior distribution for ψ; see Severini (1999), Chang and Mukerjee (2006), Chang *et al.* (2009), Ventura *et al.* (2009), Racugno *et al.* (2010) and references therein. In particular, treating $L_{mp}(\psi)$ as a genuine likelihood, the posterior distribution

$$\pi_{mp}(\psi|y) \propto \pi(\psi)\, L_{mp}(\psi) \tag{6}$$

can be obtained, where $\pi(\psi)$ is the matching prior (2) (Ventura *et al.*, 2009).

Integration of (6) over the parameter space gives the derivation of a tail area approximation, following standard Bayesian expansions (see, e.g., Reid, 2003, Brazzale *et al.*, 2007, Chapter 8). In particular, for (6) it can be shown that (Ventura and Racugno, 2010)

$$\int_{-\infty}^{\psi_0} \pi_{mp}(\psi|y)\, d\psi \quad \doteq \quad \Phi(r_p^*) \ , \tag{7}$$

where $\Phi(\cdot)$ is the standard normal distribution function and $r_p^*(\psi)$ is the modified signed likelihood ratio statistic

$$r_p^*(\psi) = r_p(\psi) + \frac{1}{r_p(\psi)} \log \frac{q(\psi)}{r_p(\psi)} \ , \tag{8}$$

with

$$r_p(\psi) = \text{sign}(\hat{\psi} - \psi)\sqrt{2(\ell_p(\hat{\psi}) - \ell_p(\psi))}$$

signed likelihood ratio statistic, $\ell_p(\psi) = \log L_p(\psi)$, and

$$q(\psi) = \frac{\ell_p'(\psi)}{|j_p(\hat{\psi})|^{1/2}} \frac{|i_{\psi\psi.\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}}{|i_{\psi\psi.\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}} \frac{|\ell_{\lambda;\hat{\lambda}}(\psi, \hat{\lambda}_\psi)|}{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2} |j_{\lambda\lambda}(\hat{\psi}, \hat{\lambda})|^{1/2}} \ , \tag{9}$$

where $j_p(\psi)$ is the profile observed information and $\ell_p'(\psi) = \partial\ell_p(\psi)/\partial\psi$. The statistic (8), with $q(\psi)$ given in (9), corresponds to the expression derived in Barndorff-Nielsen and Chamberlin (1994). In view of this, since a frequenstit p-value coincides with a Bayesian tail area probability, the prior (2) is a strong matching prior (Fraser and Reid, 2002).

The modified signed likelihood ratio statistic $r_p^*(\psi)$ is a higher-order pivotal quantity, which allows to obtain frequentist p-values, confidence limits and accurate point estimators for ψ (see, e.g., Barndorff-Nielsen and Cox, 1994, Chapter 6). In particular, an accurate confidence interval for ψ with approximate level $(1-\alpha)$ based on $r_p^*(\psi)$ is

$$\left\{\psi : |r_p^*(\psi)| \leq z_{1-\alpha/2}\right\} \ , \tag{10}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. In view of (7), (10) is also a highest posterior density credible set for ψ based on $\pi_{mp}(\psi|y)$ (Ventura and Racugno, 2010). Moreover, the modified signed likelihood ratio statistic $r_p^*(\psi)$ can also be used to derive a point estimator for ψ defined as a zero-level confidence interval, as explained in Skovgaard (1989). More precisely, the modified signed likelihood ratio statistic (8) gives rise to a simple estimating equation of the form $r_p^*(\psi) = 0$ (Pace and Salvan, 1999, Giummolé and Ventura, 2002). The corresponding estimator $\hat{\psi}^*$ is a refinement of $\hat{\psi}$, that improves its small sample properties, respecting the requirement of parameterisation equivariance. with the estimating equation $r_p^*(\psi) = 0$ giving implicitly a higher-order correction to the maximum likelihood estimator.

Finally, note that the adjustment term of the modified directed likelihood $r_p^*(\psi)$ can be decomposed into two parts, with different interpretations (see Barndorff-Nielsen and Cox, 1994, Section 6.6). In particular, $r_p^*(\psi)$ can be expressed as

$$r_p^*(\psi) = r_p(\psi) + \text{NP} + \text{INF} \ , \tag{11}$$

where NP is the nuisance parameters adjustment

$$\mathrm{NP} = -\frac{1}{r_p(\psi)} \log C(\psi) \tag{12}$$

and INF is the information adjustment

$$\mathrm{INF} = \frac{1}{r_p(\psi)} \log \frac{q(\psi)C(\psi)}{r_p(\psi)} \ . \tag{13}$$

The NP part adjusts for estimating the nuisance parameters, whereas the INF part improves the standard normal approximation. Pierce and Peters (1992) and Sartori *et al.* (1999) indicate that the NP adjustment is often appreciable, and it may yield a more substantial effect than the INF adjustment. In addition, let $r_{mp}(\psi) = \mathrm{sgn}(\hat{\psi}_{mp} - \psi)[2(\ell_{mp}(\hat{\psi}_{mp}) - \ell_{mp}(\psi))]^{1/2}$, be the signed likelihood ratio statistic based on the modified profile likelihood $L_{mp}(\psi)$, with $\ell_{mp}(\psi) = \log L_{mp}(\psi)$ and $\hat{\psi}_{mp}$ maximiser of $L_{mp}(\psi)$. Then, Sartori *et al.* (1999) show that $r_{mp}(\psi)$ is standard normal only to first order and that $r_{mp}(\psi) = r_p(\psi) + \mathrm{NP} + O(n^{-1})$. This means that in those instances where most of the asymptotic error comes from the estimation of the nuisance parameters, $r_{mp}(\psi)$ may represent a step ahead over $r_p(\psi)$ (see for example DiCiccio and Martin, 1993, DiCiccio and Stern, 1994).

3 A new modified profile likelihood

The agreement in (7) suggest to modify the profile likelihood as in (3) to define $L^*_{mp}(\psi)$. In this section we study the properties of $L^*_{mp}(\psi)$. In particular, we show that $\ell^*_{mp}(\psi) = \log L^*_{mp}(\psi)$ can be written, to second asymptotic order, as a function of $r^*_p(\psi)$, i.e.

$$\ell^*_{mp}(\psi) = -\frac{1}{2}(r^*_p(\psi))^2 + O(n^{-1}) \ ,$$

with $r^*_p(\psi)$ given in (8).

Using results in Sartori *et al.* (1999), we have that, ignoring additive constants,

$$\begin{aligned}
\ell^*_{mp}(\psi) &= \ell_{mp}(\psi) + \log \pi(\psi) \\
&= -\frac{1}{2}(r_{mp}(\psi))^2 + \log \pi(\psi) \\
&= -\frac{1}{2}(r_p(\psi))^2 - r_p(\psi)\,\mathrm{NP} + \log \pi(\psi) \\
&= -\frac{1}{2}(r_p(\psi))^2 - r_p(\psi)\left[\mathrm{NP} + \frac{1}{r_p(\psi)} \log \frac{1}{\pi(\psi)}\right] \\
&= -\frac{1}{2}(r_p(\psi))^2 - r_p(\psi)\,[\mathrm{NP} + \mathrm{INF}^*]
\end{aligned} \tag{14}$$

with NP given in (12) and $\mathrm{INF}^* = -r_p(\psi)^{-1} \log \pi(\psi)$. Since $r_p(\psi) = \ell'_p(\psi)/j_p(\hat{\psi})^{1/2} + o_p(1)$ and $\pi(\psi) \propto i_{\psi\psi.\lambda}(\psi, \hat{\lambda}_\psi)^{1/2}$, we have that $\mathrm{INF}{=}\mathrm{INF}^* + O_p(n^{-1})$ and thus

$$\begin{aligned}
\ell^*_{mp}(\psi) &= -\frac{1}{2}(r_p(\psi) + \mathrm{NP} + \mathrm{INF})^2 + O(n^{-1}) \\
&= -\frac{1}{2}(r^*_p(\psi))^2 + O(n^{-1}) \ .
\end{aligned} \tag{15}$$

This shows that $\ell^*_{mp}(\psi)$ is equal, to second asymptotic order, to a r^*-type statistics, and the quantity $\pi(\psi)$ can thus be interpreted as a further adjustment to the profile likelihood to accommodate for deviations from normality due to a small amount of information available in the sample.

In view of expansion (15), for the proposed modified profile likelihood (1) we have that the associated signed likelihood ratio statistic

$$r^*_{mp}(\psi) = \mathrm{sgn}(\hat{\psi}^*_{mp} - \psi)[2(\ell^*_{mp}(\hat{\psi}^*_{mp}) - \ell^*_{mp}(\psi))]^{1/2} ,$$

with $\hat{\psi}^*_{mp}$ maximizer of $\ell^*_{mp}(\psi)$, corresponds to $r^*_p(\psi)$ and thus is standard normal to second order. Moreover, the maximizer of $L^*_{mp}(\psi)$ can be computed as the solution of the estimating equation $r^*_p(\psi) = 0$, and thus corresponds to $\hat{\psi}^*$.

4 Examples and numerical illustrations

The aim of this Section is to provide simulation studies and an illustration of the proposed modified profile likelihood in the context of group models and exponential families. In particular, it is shown that the modified profile likelihood $L^*_{mp}(\psi)$ may represent a step ahead over $L_{mp}(\psi)$.

Example 1: Exponential family. Consider a sample $y = (y_1, \ldots, y_n)$ of n independent and identically distributed observations from an exponential family model

$$p(y; \psi, \lambda) = \exp\{\psi\, t(y) + \lambda^{\mathsf{T}} s(y) - K(\psi, \lambda)\}\, h(y) ,$$

where $t(\cdot)$, $s(\cdot)$, $h(\cdot)$ and $K(\cdot)$ are given functions. Let $(t, s) = (\sum t(y_i), \sum s(y_i))$ be the minimal sufficient statistic, with t and s associated with ψ and λ, respectively. Moreover, we denote with $K_\lambda(\psi, \lambda) = \partial K(\psi, \lambda)/\partial\lambda$, $K_\psi(\psi, \lambda) = \partial K(\psi, \lambda)/\partial\psi$, $K_{\lambda\lambda}(\psi, \lambda) = \partial^2 K(\psi, \lambda)/(\partial\lambda\partial\lambda^{\mathsf{T}})$, and so on, the partial derivatives of $K(\psi, \lambda)$.

The profile likelihood function for ψ is $L_p(\psi) = \exp\{\psi\, t + \hat{\lambda}^{\mathsf{T}}_\psi s - nK(\psi, \hat{\lambda}_\psi)\}$, where $\hat{\lambda}_\psi$ is the solution of the likelihood equation $s = nK_\lambda(\psi, \hat{\lambda}_\psi)$. It is easy to show that the modified profile likelihood of Barndorff-Nielsen (1983) reduces to

$$L_{mp}(\psi) = L_p(\psi)|K_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2} ,$$

and that the proposed modified profile likelihood (1) is

$$L^*_{mp}(\psi) = L_p(\psi)|K_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2} j_p(\psi)^{1/2} ,$$

with $j_p(\psi) = i_{\psi\psi.\lambda}(\psi, \hat{\lambda}_\psi) = nK_{\psi\psi}(\psi, \hat{\lambda}_\psi) - nK_{\psi\lambda}(\psi, \hat{\lambda}_\psi)K_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)^{-1}K_{\lambda\psi}(\psi, \hat{\lambda}_\psi)$.

Example 2: Gamma distribution. Consider a random sample $(y_1, \ldots, y_n)$ from the gamma density $p(y; \psi, \lambda) = \lambda^\psi y^{\psi-1}\exp(-\lambda y)\Gamma(\psi)^{-1}$, $y > 0$, $\psi, \lambda > 0$. We take the parameter of interest as the shape parameter ψ, with the scale parameter λ as nuisance. The profile loglikelihood is $\ell_p(\psi) = \psi(t - n) - n\log\Gamma(\psi) + n\psi\log(\psi/\bar{y})$, with $t = \sum \log y_i$ and $\bar{y}$ sample mean, and the modified profile loglikelihood (4) is $\ell_{mp}(\psi) = \ell_p(\psi) - 0.5\log\psi$. Simple calculations give $i_{\psi\psi.\lambda}(\psi, \lambda) = (n/\psi)(\rho(\psi) - 1)$,

	$n=5$	$n=10$	$n=15$	$n=20$
r_p	0.899	0.931	0.933	0.942
r_{mp}	0.937	0.944	0.944	0.947
r^*_{mp}	0.944	0.948	0.949	0.950
r^*_p	0.946	0.951	0.949	0.949

Table 1: Emprirical coverages of 0.95% confidence intervals under the gamma model.

	$n=5$	$n=10$	$n=20$
$\hat{\psi}$	1.21 (4.77)	0.35 (0.81)	0.15 (0.37)
$\hat{\psi}_{mp}$	0.82 (3.82)	0.24 (0.73)	0.09 (0.35)
$\hat{\psi}^*_{mp}$	0.03 (1.91)	0.01 (0.56)	0.002 (0.31)

Table 2: Bias (and standard deviations) of the maximizers of $\ell_p(\psi)$, $\ell_{mp}(\psi)$ and $\ell^*_{mp}(\psi)$, under the gamma model.

with $\rho(\psi) = (\partial^2/\partial\psi^2)\log\Gamma(\psi)$, and thus the proposed modified profile loglikelihood is

$$\ell^*_{mp}(\psi) = \ell_p(\psi) - \log\psi + \frac{1}{2}\log(\psi\rho(\psi) - 1)\ .$$

The behaviour of $\ell^*_{mp}(\psi)$ under the gamma model is illustrated through a simulation study based on 10000 Monte Carlo trials. Table 1 gives the empirical coverages for equi-tailed 95% confidence intervals from $r_p(\psi)$, $r_{mp}(\psi)$, $r^*_{mp}(\psi)$ and from the $r^*_p(\psi)$ statistic of Barndorff-Nielsen (1991), with $q(\psi)$ given by

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

From Table 1 we observe that, even for small n, $r^*_{mp}(\psi)$ improves on $r_{mp}(\psi)$ and exhibits a similar behaviour than $r^*_p(\psi)$. Larger sample sizes show, as one would expect, rather little differences between the results of all the procedures.

Table 2 evaluates the finite-sample properties of the maximizers of $\ell_p(\psi)$, $\ell_{mp}(\psi)$ and $\ell^*_{mp}(\psi)$. The estimators are compared in terms of the usual centering and dispersion measures, i.e. bias and standard deviation. From Table 2 it can be noted that the maximizer of $\ell^*_{mp}(\psi)$ is always preferable to $\hat{\psi}$ and $\hat{\psi}_{mp}$. This result is due to the fact that the maximizer of $\ell^*_{mp}(\psi)$ is a r^*-based estimator.

Example 3: Group model. Consider a sample $y = (y_1,\ldots,y_n)$ of n independent and identically distributed observations from a group model, with density $p(y;k,g) = p_0(g(y);k)$, where g is the group element, k the index parameter and $p_0(\cdot)$ is a given density. We also assume that k is the parameter of interest, while g is the nuisance parameter. A well-known example of group model is the composite location and scale model, with k shape parameter and $g = (\mu,\sigma)$ location-scale parameters.

The modified profile likelihood of Barndorff-Nielsen (1983) for the index parameter k can be expressed as (see Barndorff-Nielsen and Jupp, 1988)

$$L_{mp}(k) = h(\hat{g}_k)|j_{gg}(k, \hat{g}_k)|^{-1/2} L_p(k) \ ,$$

where $h(g)$ is the right invariant Haar measure on the group of transformations. Using results in Datta and Ghosh (1995), it can be shown that the adjustment needed to compute $L^*_{mp}(k)$ can be written as

$$\pi(k) \propto i_{kk.g}(k, e)^{1/2} \ , \tag{16}$$

with $i_{kk.g}(k, e) = i_{kk}(k, e) - i_{kg}(k, e) i_{gg}(k, e)^{-1} i_{gk}(k, e)$, where e is the group identity. Note that $\pi(k)$ does not depend on g, and thus on the data through $\hat{g}_k$. Note also that (16) coincides with the marginal prior density which asymptotically maximizes the expected Kullback-Leibler divergence between the marginal posterior and the prior density functions, given in Datta and Ghosh (1995).

The modified profile likelihood (3) reduces to

$$L^*_{mp}(k) = h(\hat{g}_k)|j_{gg}(k, \hat{g}_k)|^{-1/2} L_p(k)\, i_{kk.g}(k, e)^{1/2} \ .$$

Example 4: Inverse Gaussian distribution. Consider a random sample $(y_1, \ldots, y_n)$ from the inverse Gaussian distribution with density

$$p(y; \psi, \lambda) = [\psi/(2\pi y^3)]^{1/2} \exp(-\psi(y-\lambda)^2/(2\lambda^2 y)) \ ,$$

$y > 0$, $\psi, \lambda > 0$. The parameter of interest ψ is a scale parameter. The profile loglikelihood is $\ell_p(\psi) = (n/2)\log\psi - (\psi t)/(2\bar{y}^2)$, with $t = \sum(y_i - \bar{y})^2/y_i$, and the modified profile loglikelihood (4) is $\ell_{mp}(\psi) = ((n-1)/2)\log\psi - (\psi t)/(2\bar{y}^2)$. Simple calculations give $i_{\psi\psi.\lambda}(\psi, \lambda) = n/(2\psi^2)$, and thus the proposed modified profile loglikelihood is

$$\ell^*_{mp}(\psi) = \frac{(n-3)}{2}\log\psi - \frac{\psi t}{2\bar{y}^2} \ .$$

As in Example 2, the behaviour of $\ell^*_{mp}(\psi)$ under the inverse Gaussian distribution is illustrated through a simulation study based on 10000 Monte Carlo trials. Table 3 gives the empirical coverages for 95% confidence intervals from $r_p(\psi)$, $r_{mp}(\psi)$, $r^*_{mp}(\psi)$ and $r^*_p(\psi)$. From Table 3 we observe that, also in this case, $r^*_{mp}(\psi)$ improves on $r_{mp}(\psi)$ and exhibits a similar behaviour than $r^*_p(\psi)$.

Table 4 summarizes the finite-sample properties of the maximizers of $\ell_p(\psi)$, $\ell_{mp}(\psi)$ and $\ell^*_{mp}(\psi)$. From Table 4 it can be noted that, as in Example 2, the maximizer of $\ell^*_{mp}(\psi)$ always performs better than $\hat{\psi}$ and $\hat{\psi}_{mp}$.

Example 5: Skew-normal distribution. Let us consider the scalar skew-normal model (Azzalini, 1985) with density function $p(y; \psi, \mu, \sigma) = (2/\sigma)\phi((y-\mu)/\sigma)\Phi(\psi(y-\mu)/\sigma)$, $y \in \mathbb{R}$, where $\phi(x)$ denote the $N(0,1)$ density. Let the shape parameter ψ be the parameter of interest and let $\lambda = (\mu, \sigma)$, with μ and σ unknown location and scale parameters respectively, be the nuisance parameter. Estimation of the shape

	$n=5$	$n=10$	$n=15$	$n=20$
r_p	0.894	0.924	0.934	0.939
r_{mp}	0.939	0.944	0.947	0.948
r^*_{mp}	0.944	0.945	0.949	0.949
r^*_p	0.948	0.948	0.950	0.949

Table 3: Emprirical coverages of 0.95% confidence intervals under the inverse Gaussian model.

	$n=5$	$n=10$	$n=15$	$n=20$
$\hat{\psi}$	1.966 (6.77)	0.548 (1.14)	0.337 (0.69)	0.239 (0.54)
$\hat{\psi}_{mp}$	1.373 (5.42)	0.393 (1.02)	0.248 (0.64)	0.177 (0.52)
$\hat{\psi}^*_{mp}$	0.186 (2.71)	0.083 (0.79)	0.069 (0.55)	0.053 (0.46)

Table 4: Bias (and standard deviations) of the maximizers of $\ell_p(\psi)$, $\ell_{mp}(\psi)$ and $\ell^*_{mp}(\psi)$, under the inverse Gaussian model.

parameter ψ is a quite challenging problem since the profile likelihood function, as well as the modified profile likelihood (4), can be monotone increasing, giving an infinite maximum likelihood estimate. Some recent solutions, both in the frequentist and Bayesian settings, are Sartori (2006), which suggests to madify the score function for ψ, and Liseo and Loperfido (2006) and Cabras *et al.* (2010), which show, respectively, that the Jeffreys prior and the matching prior (2) for ψ are proper. In addition, Cabras *et al.* (2010) give the expressions of the modified profile likelihood $L_{mp}(\psi)$ and of $\pi(\psi)$, which is shown to be independent on λ in view of (16).

We illustrate our proposal with a quite challenging data set for the estimation of ψ. In particular, consider the *Frontier* data set, available at the package `sn` of the `R` software, which is a random sample of size $n = 50$ from a skew-normal model, with $\mu = 0$, $\sigma = 1$ and $\psi = 5$. This dataset has some interest and has been analyzed in several papers since it leads to an infinite $\hat{\psi}$, with both $L_p(\psi)$ and $L_{mp}(\psi)$ monotone functions in ψ. Sartori (2006) obtains a modified maximum likelihood estimate equal to 6.24, and the maximum likelihood estimate from $L^*_{mp}(\psi)$ is $\hat{\psi}^* = 6.3$. Figure 1 shows the modified profile likelihoods $L_{mp}(\psi)$ for ψ, which is monotone, and $L^*_{mp}(\psi)$. Our procedure gives (2.05,44.48) as a 0.95 confidence interval based on $r^*_{mp}(\psi)$.

5 Discussion

We propose an adjustment of the modified profile likelihood based on a matching prior for the parameter of interest only. By construction, the proposed modified profile likelihood has the corresponding signed likelihood ratio statistic standard normal to second order. A similar result has been discussed by Severini (2010) for an integrated likelihood for frequentist inference.

Moreover, the maximizer of the proposed modified profile likelihood is a refinement of the maximum likelihood estimator, which improves its small sample prop-

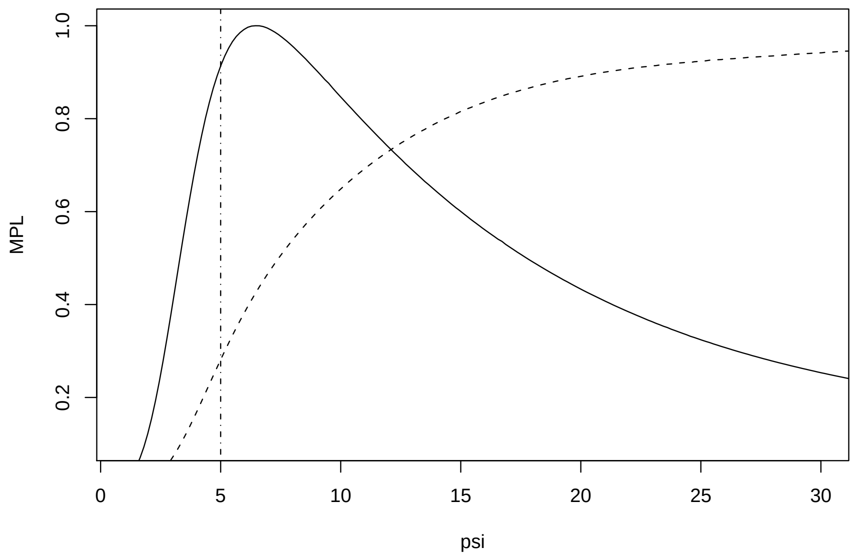

Figure 1: *Frontier* data: Plot of normalized $L_{mp}(\psi)$ (dashed) and $L^*_{mp}(\psi)$ (solid).

erties. In addition, also the results of the simulation studies for $L^*_{mp}(\psi)$ are quite good. Other results in favour of $L^*_{mp}(\psi)$, even if from a Bayesian perspective, can be found in Ventura and Racugno (2010) in the context of inference on the reliability of a stress-strength model.

As a final remark, we note that a difficulty in the application of $L^*_{mp}(\psi)$ could be the computation of the sample space derivatives involved in (5). However, in these situations, $\ell_{\lambda;\hat{\lambda}}(\psi, \hat{\lambda}_\psi)$ can be replaced in $C(\psi)$ by the approximation developed in Severini (1998), and thus using $C(\psi) = |j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}/|I(\psi, \hat{\lambda}; \hat{\psi}, \hat{\lambda})|$, where $I(\psi, \lambda; \psi_0, \lambda_0) = E_{(\psi_0,\lambda_0)}(\ell_\lambda(\psi, \lambda)\ell_\lambda(\psi_0, \lambda_0)^\top)$, with $\ell_\lambda(\psi, \lambda) = \partial\ell(\psi, \lambda)/\partial\lambda$, in $L^*_{mp}(\psi)$. In this case, also in the expression of $q(\psi)$ given in (9) the sample space derivatives have to be replaced using Severini's approximation.

References

[1] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.*, **12**, 171-178.

[2] Barndorff-Nielsen, O.E, (1980). Conditionality resolutions. *Biometrika*, **67**, 293-310.

[3] Barndorff-Nielsen, O.E, (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.

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

[5] Barndorff-Nielsen, O.E, Chamberlin, S.R. (1994). Stable and invariant adjusted directed likelihoods. *Biometrika*, **81**, 485–499.

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

[7] Barndorff-Nielsen, O. E. and Jupp, P. E. (1988). Differential geometry, profile likelihood, L-sufficiency and composite transformation models. *Ann. Statist.*, **16**, 1009-1043.

[8] Berger, J.O., Liseo, B., Wolpert, R.L. (1999). Integrated likelihood methods for eliminating nuisance parameters. *Statist. Sci.*, **14**, 1-22.

[9] Brazzale, A. R., Davison, A. C., Reid, N. (2007). *Applied Asymptotics: Case Studies in Small Sample Statistics.* Cambridge University Press, Cambridge.

[10] Cabras, S., Castellanos, M.E., Racugno, W., Ventura, L. (2010). A matching prior for the shape parameter of the skew-normal distribution. Under revision.

[11] Chang, H., Mukerjee, R. (2006). Probability matching property of adjusted likelihoods. *Statist. Probab. Lett.*, **76**, 838-842.

[12] Chang, H., Kim, B.H., Mukerjee, R. (2009). Bayesian and frequentist confidence intervals via adjusted likelihoods under prior specification on the interest parameter. *Statistics*, **43**, 203-211.

[13] Datta, G.S., Ghosh, J.K. (1995). Noninformative priors for maximal invariant parameter in group models. *Test*, **4**, 95-114.

[14] Datta, G. S., Mukerjee, R. (2004). *Probability Matching Priors: Higher Order Asymptotics.* Springer, Berlin.

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

[16] Diciccio, T.J., Stern, S.E. (1994). Constructing approximately standard normal pivots from signed roots of adjusted likelihood ratio statistics. *Scand. J. Statist.*, **21**, 447-460.

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

[18] Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika*, **80**, 3-26.

[19] Fraser, D.A.S. (2003). Likelihood for component parameters. *Biometrika*, **90**, 327-339.

[20] Fraser, D.A.S., Reid, N. (2002). Strong matching of frequentist and Bayesian parametric inference. *J. Statist. Plan. Inf.*, **103**, 263–285.

[21] Giummolé, F., Ventura, L. (2002). Practical point estimation from higher-order pivots. *J. Statist. Comput. Simul.*, **72**, 419-430.

[22] Greco, L., Racugno, W., Ventura, L. (2008). Robust likelihood functions in Bayesian inference. *J. Statist. Plann. Inf.*, **138**, 1258-1270.

[23] Lazar, N.A. (2003). Bayesian empirical likelihood. *Biometrika*, **90**, 319-326.

[24] Lin, L. (2006). Quasi Bayesian Likelihood. *Statistical Methodology*, **3**, 444455.

[25] Liseo, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika*, **80**, 295-304.

[26] Liseo, B., Loperfido, N. (2006). A note on reference priors for the scalar skew-normal distribution. *J. Statist. Plann. Inf.*, **136**, 373-389.

[27] McCullagh, P. Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *J. Roy. Statist. Soc.* B, **52**, 325-344.

[28] Mukerjee, R., Reid, N. (2000). On the Bayesian approach for frequentist computations. *Braz. J. Prob. Statist.*, **14**, 159-166.

[29] Pace, L., Salvan, A. (1999). Point estimation based on confidence intervals: Exponential families. *J. Statist. Comput. Simul.*, **64**, 1-21.

[30] Pace, L., Salvan, A. (2006). Adjustments of the profile likelihood from a new perspective. *J. Statist. Plan. Inf.*, **136**, 3554–3564.

[31] Pauli, F., Racugno, W., Ventura, L. (2011). Bayesian composite marginal likelihoods. *Statistica Sinica*, **21**, 149-164.

[32] Pierce, D.A., Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families. *J. Roy. Statist. Soc.* B, **54**, 701-725.

[33] Racugno, W., Salvan, A., Ventura, L. (2010). Bayesian analysis in regression models using pseudo-likelihoods. *Comm. Stat. Th. Meth.*, **39**, 3444 3455.

[34] Reid, N. (1995). Likelihood and Bayesian approximation methods. *Bayesian Statistics*, **5**, 351-368.

[35] Sartori, N. (2006). Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *J. Statist. Plann. Inf.*, **136**, 42594275.

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

[37] Severini, T.A. (1998). Likelihood functions for inference in the presence of a nuisance parameter. *Biometrika*, **85**, 507-522.

[38] Severini, T.A. (1999). On the relationship between Bayesian and non-Bayesian elimination of nuisance parameters. *Stat. Sin.*, **9**, 713-724.

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

[40] Severini, T.A. (2007). Integrated likelihood functions for non-Bayesian inference. *Biometrika*, **94**, 529-542.

[41] Severini, T.A. (2010). Likelihood ratio statistics based on an integrated likelihood. *Biometrika*, **97**, 116.

[42] Severini, T.A. (2011). Frequency properties of inferences based on an integrated likelihood function. *Statistica Sinica*, **21**, 433-447.

[43] Stern, S.E. (1997). A second-order adjustment to the profile likelihood in the case of a multidimensional parameter of interest. *J. Roy. Statist. Soc.* B, **59**, 653-665.

[44] Ventura, L., Cabras, S., Racugno, W. (2009). Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. *J. Amer. Stat. Assoc.*, **104**, 768–774.

[45] Ventura, L., Cabras, S., Racugno, W. (2010). Default prior distributions from quasi- and quasi-profile likelihoods. *J. Statist. Plann. Inf.*, **140**, 2937-2942.

[46] Ventura, L., Racugno, W. (2010). Recent advances on Bayesian inference for $P(X < Y)$. *Working Papers*, 2010.7, Dipartimento di Scienze Statistiche, Università degli Studi di Padova.