

BIBLIOTECA DI SCIENZE STATISTICHE

SERVIZIO BIBLIOTECARIO NAZIONALE

BID P010850385 BID

ACQ. 485 / '02 INV. 82185

COLL. CLASS. 5-Coll. WP. 2002/10

**Modifications to the profile
likelihood in models with
incidental nuisance
parameters**

N. Sartori

2002.10

UNIVERSITA' DI PADOVA
BIBLIOTECA DI SCIENZE STATISTICHE
Via C. Battisti, 241 - 35121 PADOVA

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

luglio 2002

Modifications to the profile likelihood in models with incidental nuisance parameters

By N. SARTORI

Department of Statistics,

University of Padova, via C. Battisti 241-243, 35121 Padova, Italy

sartori@stat.unipd.it

SUMMARY

We study asymptotic properties of the profile and modified profile likelihoods in models for clustered data with incidental nuisance parameters. To this end, we use a two-index asymptotics setting. This means that both the sample size of the clusters, m , and the dimension of the nuisance parameter, q , may increase to infinity. It is shown that modified profile likelihoods give improvements, with respect to the profile likelihood, on consistency of estimates and on asymptotic distribution properties. In particular, the profile likelihood based statistics have the usual asymptotic distribution, provided that $1/m = o(q^{-1})$, while the analogous condition for modified profile likelihoods is $1/m = o(q^{-1/3})$.

Some key words: Conditional likelihood; Incidental nuisance parameter; Modified profile likelihoods; Profile likelihood; Profile score bias; Two-index asymptotics.

1. INTRODUCTION

Consider a model with a parameter θ which may be written in the form $\theta = (\psi, \lambda)$, where ψ is a parameter of interest and λ is a nuisance parameter. In this paper we will study asymptotic properties of profile and modified profile likelihoods in models with incidental nuisance parameters, that means that the dimension of λ increases with the sample size n . In this context, ψ is also called structural parameter.

Likelihood inference for ψ is usually based on the profile likelihood, which is the likelihood with the nuisance parameter replaced by its maximum likelihood estimate for fixed ψ . In the usual setting, with independent observations and fixed dimension of θ , the profile likelihood has some, but not all, properties of a proper likelihood. In particular, the expected value of the profile score is not zero. This defect is of primary interest here.

A modified profile likelihood for ψ may be seen, as the name suggests, as a modification to the profile likelihood, even though the modification can be obtained from different perspectives. In this paper we consider well known examples, such as those of Barndorff-Nielsen (1983, 1994, 1995), Cox & Reid (1987), McCullagh & Tibshirani (1990), Severini (1998), among others. Typically, the modification is such that the expected value of the modified profile score is closer to zero than that of the profile.

The reduced score bias is not reflected in improved formal properties, in the usual asymptotic setting with fixed dimension of the nuisance parameter. In fact, the likelihood ratio statistic has first-order asymptotic chi-squared distribution for both profile and modified profile likelihoods, and the rate of convergence of the maximum likelihood estimator to the true parameter value ψ is in both cases $n^{-1/2}$. However, when $\dim(\lambda)$ is large, the distribution properties of the modified profile likelihood are often better, for given n , than those

of the profile likelihood. See for example DiCiccio & Stern (1994) and Sartori *et al.* (1999). The failure of standard asymptotics to distinguish between the behavior of profile and modified profile likelihood when $\dim(\lambda)$ is large can be rectified by considering asymptotics where $\dim(\lambda)$ increases with n .

The study of models where the dimension of the parameter may increase with the sample size has been previously analyzed also in Portnoy (1988), Pierce & Peters (1992, § 3), Barndorff-Nielsen & Cox (1994, § 8.5) and Barndorff-Nielsen (1996).

In the following, we will refer to a two-index asymptotic setting for independent clustered observations of the form

$$Y_{ij} \sim p_{ij}(y_{ij}; \psi, \lambda_i),$$

where $i = 1, \dots, q$ and $j = 1, \dots, m_i$. The sample size is $n = \sum_{i=1}^q m_i$ and the nuisance parameter $\lambda = (\lambda_1, \dots, \lambda_q)$ has dimension q . We stressed the dependence of the model function on i and j to include situations with explanatory variables x_{ij} . In such cases, we can write $p_{ij}(y_{ij}; \psi, \lambda_i) = p(y_{ij}; \psi, \lambda_i, x_{ij})$. However, the dependence of p on i and j will be dropped from notation in the following. The two-index asymptotics setting will be called $m \times q$ -asymptotics, as defined in Barndorff-Nielsen (1996) and also suggested in Davison (1992). This means that both number of clusters, q , and cluster sample sizes, m_i , are allowed to increase to infinity. We will assume for simplicity $m_i = m$. But we could alternatively assume that each m_i can be written in the form $m_i = K_i m$, with $A \leq K_i \leq B$ and where A and B are positive finite numbers.

The sample size is $n = m q$. This means that, if we fix q , we have a nuisance parameter of fixed dimension and the standard asymptotic theory applies. On the contrary, if m is fixed, we have $O(n) = O(q)$ and maximum likelihood estimators are likely to be inconsistent, as is well known since Neyman & Scott (1948). In some cases, it is pos-

sible to solve this problem using some inferential separation in the likelihood, as with conditional or marginal likelihoods. One notable example is the conditional likelihood for components of the canonical parameter in a full exponential family (Andersen, 1970, 1971). However, conditional or marginal likelihoods are available only when the model has a particular structure, while profile and modified profile likelihoods are general tools for inference. With $m \times q$ -asymptotics we can emulate the case with m fixed, by considering m that increases slowly with q .

The main result of this paper is that, while methods based on the profile likelihood fail, unless $1/m = o(q^{-1})$, methods based on modified profile likelihoods may still perform accurately, provided that $1/m = o(q^{-1/3})$. The result on the profile likelihood was already known in the context of regular exponential families (Portnoy, 1988), although expressed in a different form. Indeed, since $n = mq$, the above conditions can be written equivalently as $q = o(n^{1/2})$ and $q = o(n^{3/4})$, respectively.

The outline of the paper is as follows. Section 2 gives some notation and preliminaries. Section 3 shows the different $m \times q$ -asymptotic properties of the profile and modified profile score statistics. Section 4 deals with related likelihood quantities. Section 5 contains examples and simulations. Finally, a brief discussion is given in Section 6.

2. NOTATION AND PRELIMINARIES

As introduced in the previous section, we will consider models with independent clustered observations

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

where $i = 1, \dots, q$ and $j = 1, \dots, m$. We assume for simplicity that ψ and λ_i are scalars, although this assumption is not crucial for the results in the paper.

The loglikelihood can be written as

$$l(\theta) = \sum_{i=1}^q l^i(\psi, \lambda_i), \quad (1)$$

where

$$l^i(\psi, \lambda_i) = \sum_{j=1}^m \log p(y_{ij}; \psi, \lambda_i)$$

is the loglikelihood function relative to the i -th cluster. We will assume usual regularity conditions for the loglikelihoods $l^i(\psi, \lambda_i)$ (see, for instance, § 3.4 of Severini, 2000). Further regularity conditions will be stated below.

The maximum likelihood estimator of $\theta = (\psi, \lambda)$ will be denoted by $\hat{\theta} = (\hat{\psi}, \hat{\lambda})$. The score function will be denoted by $U(\theta) = \partial l(\theta) / \partial \theta$, while the observed and expected information will be indicated by $j(\theta)$ and $i(\theta)$, respectively. For single components we will use subscripts. Moreover, in some cases, arguments will be dropped from notation to avoid messy expressions. As an example, we will use $U_\psi = U_\psi(\theta) = \partial l(\theta) / \partial \psi$, $j_{\psi\lambda_i} = j_{\psi\lambda_i}(\theta) = -\partial U_\psi(\theta) / \partial \lambda_i$ and $i_{\psi\lambda_i} = i_{\psi\lambda_i}(\theta) = E_\theta(j_{\psi\lambda_i})$.

As can be seen from (1), the loglikelihood is separable with respect to nuisance parameters, that means that it is the sum of q terms, where each of them depends only on one nuisance parameter. This is due to the fact that the nuisance parameter λ_i is related just to the i -th cluster and that the clusters are independent. This implies that, for each single nuisance parameter, we can consider the standard asymptotics in powers of m . For instance, we have $U_{\lambda_i}(\theta) = U_{\lambda_i}^i(\psi, \lambda_i) = O_p(m^{1/2})$, $j_{\lambda_i\lambda_i}(\theta) = j_{\lambda_i\lambda_i}^i(\psi, \lambda_i) = O_p(m)$ and $j_{\lambda_i\lambda_k}(\theta) = 0$, when $i \neq k$. Moreover, the separability of nuisance parameters also implies that the constrained maximum likelihood estimate of λ for fixed ψ , $\hat{\lambda}_\psi$, is the solution of the q independent likelihood equations for the clusters. As a consequence, the profile loglikelihood for ψ may be written as the

sum of the q profile loglikelihoods for the clusters

$$l_P(\psi) = l(\psi, \hat{\lambda}_\psi) = \sum_{i=1}^q l^i(\psi, \hat{\lambda}_{i\psi}) = \sum_{i=1}^q l_P^i(\psi). \quad (2)$$

In the following, U_P will denote the profile score function. After standard expansions, such as those in the Appendix of McCullagh & Tibshirani (1990), we have the following expansion for the profile score in the i -th cluster

$$U_P^i = U_{\psi|\lambda_i} + B_P^i + R_P^i, \quad (3)$$

where $U_{\psi|\lambda_i} = U_\psi^i - i_{\psi\lambda_i} i_{\lambda_i\lambda_i}^{-1} U_{\lambda_i}$, B_P^i is a term of order $O_p(1)$ with expected value $-\rho_\psi^i$, of order $O(1)$, and R_P^i is the remainder term of order $O_p(m^{-1/2})$ and with expected value of order $O(m^{-1})$ (see also DiCiccio *et al.*, 1996). Hence, we have that the bias of the profile score in the i -th cluster is $E_\theta(U_P^i) = -\rho_\psi^i + O(m^{-1})$. As noted also by McCullagh & Tibshirani (1990), the problem with the profile likelihood for clustered data is that the bias of the profile score accumulates across clusters. In fact, considering (2), it is straightforward to see that the profile score bias has leading term $-\sum_{i=1}^q \rho_\psi^i = O(q)$.

In the following, we will consider modified profile loglikelihoods of the form

$$l_M(\psi) = l_P(\psi) + M(\psi),$$

where the modification term $M(\psi)$ satisfies

$$M(\psi) = \sum_{i=1}^q M^i(\psi), \quad (4)$$

and $M^i(\psi)$ is a suitably smooth function having derivatives of order $O_p(1)$ and such that

$$E_\theta \left(\frac{\partial M^i}{\partial \psi} \right) = \rho_\psi^i + O(m^{-1}). \quad (5)$$

The last property means that $E_\theta(\partial M^i / \partial \psi)$ removes all but $O(m^{-1})$ of the profile score bias of the cluster. Properties (4) and (5) are generally

satisfied by most modifications proposed in literature, such as those of McCullagh & Tibshirani (1990), Cox & Reid (1987), Barndorff-Nielsen (1994, 1995), among others. Throughout the paper, we will mainly refer to the modified profile likelihood of Barndorff-Nielsen (1983), which is a highly accurate approximation to a conditional or marginal likelihood, when either exists. Moreover, it is also invariant with respect to interest respecting reparameterizations. It has modification of the form

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

with $l_{\lambda;\hat{\lambda}}(\theta) = \partial l(\theta; \hat{\theta}, a) / (\partial \lambda \partial \hat{\lambda}^T)$. The computation of this term requires a sample space derivative. This means that we need to write the data y in the form $(\hat{\theta}, a)$, where a is an ancillary statistic, either exactly or approximately. This is straightforward in full exponential families and in transformation models. In these two class of models, the modified profile likelihood satisfies properties (4) and (5). For a general model, we can use approximations for $M(\psi)$. See Severini (2000, § 9.5) for a recent review of approximations that do not require sample space derivatives and still satisfy (4) and (5).

Example 1: Full exponential family. Let us consider a full exponential family with a component of the canonical parameter as the parameter of interest. With clustered data, the loglikelihood can be written in the form

$$l(\theta) = \sum_{i=1}^q l^i(\psi, \lambda_i) = \sum_{i=1}^q \{ \psi u_i + \lambda_i v_i - m K(\psi, \lambda_i) \},$$

where the sufficient statistic in the i -th cluster has components $u_i = \sum_j u_{ij}$ and $v_i = \sum_j v_{ij}$ and $K(\cdot)$ is the cumulant function. The overall sufficient statistic has components $u = \sum_{i=1}^q u_i$ and $v = (v_1, \dots, v_q)$.

The modified profile loglikelihood has

$$M(\psi) = \frac{1}{2} \log |K_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)| = \sum_{i=1}^q M^i(\psi) = \sum_{i=1}^q \frac{1}{2} \log K_{\lambda_i \lambda_i}(\psi, \hat{\lambda}_{i\psi}),$$

where $K_{\lambda\lambda} = \partial K / (\partial \lambda \partial \lambda^T)$, and is exactly the sum of the modified profile loglikelihoods of each single cluster. This may not be the case in general models.

3. SCORE STATISTICS

In this section we consider the $m \times q$ -asymptotic distribution of the profile and modified profile score statistics

$$\begin{aligned} W_P^u(\psi) &= j_P(\psi)^{-1} U_P(\psi)^2 \\ W_M^u(\psi) &= j_M(\psi)^{-1} U_M(\psi)^2, \end{aligned}$$

where j_P and j_M are the profile and modified profile observed information, and U_M is the modified profile score function.

Let us consider first the profile score function. Using (2) and (3), we can write

$$U_P = U_{\psi|\lambda} + B_P + R_P, \quad (6)$$

where $U_{\psi|\lambda} = U_\psi - i_{\psi\lambda} i_{\lambda\lambda}^{-1} U_\lambda = \sum_{i=1}^q U_{\psi|\lambda_i}$, $B_P = \sum_{i=1}^q B^i$ and $R_P = \sum_{i=1}^q R_P^i$. Note that the variance of $U_{\psi|\lambda}$ is $i_{\psi\psi|\lambda} = i_{\psi\psi} - i_{\psi\lambda} i_{\lambda\lambda}^{-1} i_{\lambda\psi}$. This quantity is also called the partial information for ψ (see, for instance, § 3.6.3 in Severini, 2000) and, as $U_{\psi|\lambda}$, is additive among clusters. Indeed, the independence between clusters implies that $i_{\psi\psi|\lambda} = \sum_{i=1}^q V_\theta(U_{\psi|\lambda_i}) = \sum_{i=1}^q i_{\psi\psi|\lambda_i}$.

In the following, we will assume that the sequence of quantities $\bar{i}_{\psi\psi|\lambda} = i_{\psi\psi|\lambda} / n = i_{\psi\psi|\lambda} / (mq)$ converges to a strictly positive number, as q and m diverge. This reasonable assumption allows us to write $i_{\psi\psi|\lambda} = n \bar{i}_{\psi\psi|\lambda}$ and it guarantees an asymptotic lower bound for the partial information.

It is straightforward to see that $U_{\psi|\lambda}$ is of order $O_p(n^{1/2})$. In fact, $U_{\psi|\lambda}$ is a sum of n independent quantities with zero mean and sum of the variances equal to $i_{\psi\psi|\lambda}$. Then, as n goes to infinity and subject to

general stability conditions, the quantity $Z = i_{\psi\psi|\lambda}^{-1/2} U_{\psi|\lambda}$ is asymptotically standard normal, and regardless of the nature of the sequence $\{q, m\}$. Hence $Z = O_p(1)$ and, due to the assumption on $i_{\psi\psi|\lambda}$, $U_{\psi|\lambda}$ is of order $O_p(n^{1/2})$. We note that Z cannot be used inferentially, since it depends on λ . The order of the remaining terms in (6) is evaluated analogously using the central limit theorem for independent variables. In particular, as shown in the Appendix, B_P is of order $O_p(q)$, as a consequence of the score bias, and $R_P = O_p[\max\{q/m, (q/m)^{1/2}\}]$. The rather unconventional latter expression is required by the unconventional $m \times q$ -asymptotics, with the two terms corresponding respectively to whether it is the mean or the standard deviation of R_P that dominates.

Let us consider now a modified profile likelihood such that the modification term $M(\psi)$ satisfies (4) and (5). The modified profile score function $U_M = U_P + \partial M / \partial \psi$ is

$$U_M = U_{\psi|\lambda} + B_M + R_P,$$

where $B_M = B_P + \partial M / \partial \psi$. The crucial point is that B_M is of the same order as R_P . The reason for this is that the adjustment $\partial M / \partial \psi$ removes a major part of the profile score bias, leaving only a part that is of the same order of R_P . This is due to properties (4) and (5), together with expansion (6), as shown in the Appendix. Hence we can write

$$U_M = U_{\psi|\lambda} + R_M,$$

where $R_M = B_M + R_P$ is a quantity of the same order as R_P .

In what follows we use temporarily the partial information in place of the observed information for standardization of score functions. The main result here is that the standardized profile and modified profile scores have the form

$$\begin{aligned} i_{\psi\psi|\lambda}^{-1/2} U_P &= Z + i_{\psi\psi|\lambda}^{-1/2} B_P + i_{\psi\psi|\lambda}^{-1/2} R_P \\ i_{\psi\psi|\lambda}^{-1/2} U_M &= Z + i_{\psi\psi|\lambda}^{-1/2} R_M, \end{aligned}$$

where in the latter case, the dominant term $i_{\psi\psi|\lambda}^{-1/2} B_P = O_p\{(q/m)^{1/2}\}$ has been removed by the bias correction of the modified profile likelihood. On the other hand, both residual terms, $i_{\psi\psi|\lambda}^{-1/2} R_P$ and $i_{\psi\psi|\lambda}^{-1/2} R_M$, are of order $O_p[\max\{(q/m^3)^{1/2}, m^{-1}\}] = o_p\{(q/m)^{1/2}\}$.

For the asymptotic normality of the standardized score statistics we need Z to be the leading term in the expansions above, that means that the remaining terms need to be $o_p(1)$. In the first case, we have

$$i_{\psi\psi|\lambda}^{-1/2} U_P = Z + O_p\{(q/m)^{1/2}\} = Z + O_p(q n^{-1/2}). \quad (7)$$

This implies that we need $q/m = o(1)$ for $i_{\psi\psi|\lambda}^{-1/2} U_P$ to be asymptotically standard normal. This means that we require the sample size in each cluster to increase at a faster rate than the number of clusters. The condition can be written in the equivalent forms $1/m = o(q^{-1})$ and $q = o(n^{1/2})$. On the other hand, for a modified profile likelihood, we have

$$i_{\psi\psi|\lambda}^{-1/2} U_M = Z + O_p[\max\{(q/m^3)^{1/2}, m^{-1}\}].$$

This implies that, in this case, the condition becomes $q/m^3 = o(1)$. This may be stated in different, but equivalent, forms, such as $1/m = o(q^{-1/3})$ or $q = o(n^{3/4})$. In particular, the condition $1/m = o(q^{-1/3})$ shows that, if we let q increase, m has to increase as well, and it has to increase faster than $q^{1/3}$. From another point of view, the equivalent condition $q = o(n^{3/4})$ says that q can increase with n , but it has to increase slower than $n^{3/4}$. Anyway, what is important to note is that this is a stronger result than the one for the profile likelihood. Indeed, in situations in which m increases faster than $q^{1/3}$, but not faster than q , $i_{\psi\psi|\lambda}^{-1/2} U_M$ has the usual asymptotic distribution while this cannot be guaranteed for $i_{\psi\psi|\lambda}^{-1/2} U_P$.

We note that the results do not change whether we use $i_{\psi\psi|\lambda}$ or the observed information. Indeed, in formula (13) in the Appendix, it is shown that $j_P = i_{\psi\psi|\lambda}\{1 + o_p(1)\}$. This means that

$$j_P^{-1/2} U_P = i_{\psi\psi|\lambda}^{-1/2} U_P \{1 + o_p(1)\},$$

where the relative error is of order $O_p(n^{-1/2})$ when $1/m = o(q^{-1})$ and $O_p(m^{-1})$ otherwise. Hence, also $j_P^{-1/2}U_P$ is asymptotically standard normal, provided that $1/m = o(q^{-1})$. This also implies that $W_P^u = Z^2 + o_p(1)$, giving the usual χ_1^2 asymptotic distribution. Analogously, we have

$$j_M^{-1/2}U_M = i_{\psi|\lambda}^{-1/2}U_M \{1 + o_p(1)\}.$$

Hence, $j_M^{-1/2}U_M$ is asymptotically standard normal, and W_M^u asymptotically χ_1^2 , provided that $1/m = o(q^{-1/3})$.

As a final remark, we note that, when $1/m = o(q^{-1})$, we have

$$j_M^{-1/2}U_M = Z + O_p(n^{-1/2})$$

as opposed to (7). Hence, even in situations when both profile and modified profile score statistics have the usual asymptotic distribution, the error term has a smaller upper bound in the latter case.

4. OTHER LIKELIHOOD QUANTITIES

Here, we will see how the results of the previous section reflect on consistency of estimates and on the asymptotic distribution of other likelihood based statistics, such as the likelihood ratio and the Wald statistics. In the latter case, it is shown that the three versions of likelihood based statistics are asymptotically equivalent, as in the standard asymptotic setting. For what concerns the estimators, it is shown that modified profile likelihoods give improvements in terms of consistency over the profile likelihood.

In general, denoting by $\hat{\psi}_M$ the maximizer of l_M , we have that $\hat{\psi}$ and $\hat{\psi}_M$ will be consistent, no matter what is the nature of the sequence $\{m, q\}$, since both m and q go to infinity. However, the rates of convergence to the true parameter value depend on the relative rate of q and m .

Let us consider the profile case. An expansion for the profile likelihood equation around ψ gives

$$0 = U_P(\hat{\psi}) = U_P(\psi) - j_P(\psi)(\hat{\psi} - \psi) + O_p(n)O_p\{(\hat{\psi} - \psi)^2\},$$

where we assume that third and subsequent derivatives of the profile loglikelihood are of order $O_p(n)$. This implies that

$$(\hat{\psi} - \psi) = j_P(\psi)^{-1}U_P(\psi) + O_p\{(\hat{\psi} - \psi)^2\}. \quad (8)$$

Using (6) and (13), it is straightforward to see that $\hat{\psi} = \psi + O_p(n^{-1/2})$ if $1/m = o(q^{-1})$, while $\hat{\psi} = \psi + O_p(m^{-1})$ otherwise. On the other hand, using an analogous reasoning for the modified profile likelihood, we have that $\hat{\psi}_M = \psi + O_p(n^{-1/2})$ when $1/m = o(q^{-1/3})$, and $\hat{\psi}_M = \psi + O_p(m^{-2})$ otherwise. Hence, when q increases faster than m , $\hat{\psi}_M$ might converge to ψ at a faster rate than $\hat{\psi}$. In some sense, this formalizes the comment in Barndorff-Nielsen & Cox (1994, p. 285), saying that, in $m \times q$ -asymptotics, $\hat{\psi}_M$ “will generally be more nearly consistent than $\hat{\psi}$ ”.

Let us consider now the likelihood ratio and Wald statistics based on profile and modified profile likelihoods. In particular, the likelihood ratio statistics are

$$W_P(\psi) = 2\{l_P(\hat{\psi}) - l_P(\psi)\}, \quad W_M(\psi) = 2\{l_M(\hat{\psi}_M) - l_M(\psi)\},$$

and the Wald statistics are

$$W_P^e(\psi) = j_P(\psi)(\hat{\psi} - \psi)^2, \quad W_M^e(\psi) = j_M(\psi)(\hat{\psi}_M - \psi)^2.$$

In the standard asymptotic setting, score, Wald and likelihood ratio statistics are first-order asymptotically equivalent. It is possible to see that this is still true in $m \times q$ -asymptotics. Indeed, using (8) and the above results on $(\hat{\psi} - \psi)$, it follows that

$$W_P^e = W_P^u \{1 + O_p(n^{-1/2})\},$$

provided that $1/m = o(q^{-1})$. If this condition does not hold, the above relation becomes $W_P^e = W_P^u \{1 + O_p(m^{-1})\}$. Similarly, expanding $l_P(\psi)$ around $\hat{\psi}$, we obtain

$$W_P = W_P^e \{1 + O_p(n^{-1/2})\},$$

when $1/m = o(q^{-1})$, and $W_P = W_P^e \{1 + O_p(m^{-1})\}$ otherwise. This implies that W_P and W_P^e have asymptotic χ_1^2 distribution, as W_P^u , provided that $1/m = o(q^{-1})$.

For the modified profile likelihood, we follow the same steps and we obtain

$$W_M^e = W_M^u \{1 + O_p(n^{-1/2})\}, \quad W_M = W_M^e \{1 + O_p(n^{-1/2})\},$$

when $1/m = o(q^{-1/3})$, while the relative error is of order $O_p(m^{-1})$ otherwise. This means that W_M and W_M^e have χ_1^2 asymptotic distribution, provided that $1/m = o(q^{-1/3})$, as for W_M^u .

What has been shown is that the three likelihood based statistics are equivalent. This means that when one of them has the usual asymptotic distribution, the other two are equivalent to it with a relative error of order $O_p(n^{-1/2})$, as in standard asymptotics. Also the reverse is true, i.e. when one fails also the other two fail. The conditions for the usual asymptotic results are those found in the previous section.

As a final remark, we note that, in the formulae for the score and Wald statistics, we can alternatively use the observed information evaluated at the estimate of ψ , obtaining the same asymptotic results. This is easily shown by expanding the observed information around the maximum likelihood estimate.

5. EXAMPLES

Example 1: Full exponential family (cont.). In full exponential families, a conditional likelihood for the canonical parameter ψ is available,

at least in principle, that is the likelihood associated with the conditional distribution of U given V . Hence, profile and modified profile likelihoods can be also compared with this conditional likelihood. In particular, the conditional loglikelihood is given by

$$l_C(\psi) = l(\theta) - l(\theta; v),$$

where $l(\theta; v)$ is the loglikelihood based on the marginal distribution of V . In the standard asymptotic setting, the conditional likelihood can be approximated using a saddlepoint approximation for the marginal distribution of V (see, for instance, § 8.2.4 of Severini, 2000). This gives the modified profile likelihood with an error term of order $O(n^{-1})$. In $m \times q$ -asymptotics, we can apply a similar reasoning. First of all, note that the conditional loglikelihood for ψ can be written in the form

$$l_C(\psi) = \sum_{i=1}^q l_C^i(\psi), \quad (9)$$

where $l_C^i(\psi)$ is the conditional loglikelihood obtained from the distribution of U_i given V_i . The additivity property holds for the profile and the modified profile loglikelihood as well, as already seen in Section 2. Moreover, in each cluster, it is well known that we have

$$\begin{aligned} l_C^i(\psi) &= l_P^i(\psi) + O_p(1) \\ l_C^i(\psi) &= l_P^i(\psi) + M^i(\psi) + O_p(m^{-1}). \end{aligned} \quad (10)$$

Note that $M^i(\psi)$ and its derivatives are of order $O_p(1)$, and derivatives of the residual term in (10) are also of order $O_p(m^{-1})$. This implies $l_C(\psi) = l_P(\psi) + O_p(q)$ and $l_C(\psi) = l_M(\psi) + O_p(q/m)$. However, the relation between $l_P(\psi)$, $l_M(\psi)$ and $l_C(\psi)$ is better explained when we consider the relative version of these loglikelihoods. We will denote the relative loglikelihoods with $\bar{l}_P(\psi)$, $\bar{l}_M(\psi)$ and $\bar{l}_C(\psi)$. After some

standard expansions it is possible to prove that

$$\begin{aligned}\bar{l}_P(\psi) &= \bar{l}_C(\psi) + o_p(1), \quad \text{when } 1/m = o(q^{-1}) \\ \bar{l}_M(\psi) &= \bar{l}_C(\psi) + o_p(1), \quad \text{when } 1/m = o(q^{-1/3}).\end{aligned}$$

These results generalize those of Barndorff-Nielsen (1996) regarding the gamma distribution and are in agreement with those of Section 4. In fact, the conditional likelihood ratio statistics, $W_C(\psi) = -2\bar{l}_C(\psi)$, has an asymptotic χ_1^2 distribution, as $q \rightarrow \infty$. This is true even in the case with m fixed (Andersen, 1971). The same result holds for $W_P(\psi)$, if $1/m = o(q^{-1})$, and for $W_M(\psi)$, if $1/m = o(q^{-1/3})$.

A special case is when Y_{ij} are independent normal with mean λ_i and variance ψ . Indeed, the modified profile likelihood for ψ is exactly equal to the conditional likelihood, which is also a marginal likelihood (see, for instance, Example 9.6 in Severini, 2000). The same thing happens when Y_{ij} are independent inverse Gaussian random variables; see Example 1 in Sartori *et al.* (1999). Examples 2 and 3 below are instances where the conditional likelihood exists, but it is not equal to the modified profile likelihood.

Example 2: Gamma samples with common shape parameter. Let us consider Y_{ij} as independent gamma random variables with shape parameter ψ and scale parameter $1/\lambda_i$, as in Example 5.1 of Barndorff-Nielsen (1996). The sufficient statistic has components $u = \sum_i^q \sum_j^m \log y_{ij}$ and $v_i = \sum_j^m y_{ij}$, $i = 1, \dots, q$. Writing $s = u - m \sum_i^q \log v_i$, the log-likelihoods are

$$\begin{aligned}l_C(\psi) &= \psi s + q \log \Gamma(m\psi) - m q \log \Gamma(\psi) \\ l_P(\psi) &= \psi s + m q \psi \log m\psi - m q \psi - m q \log \Gamma(\psi) \\ l_M(\psi) &= \psi s + q(m\psi - 0.5) \log m\psi - m q \psi - m q \log \Gamma(\psi),\end{aligned}$$

which have to be maximized numerically. Denoting by r_C , r_P and r_M the signed square root of the conditional, profile and modified profile

likelihood ratio statistics, Table 1 reports the probabilities $\Phi(r_P(\psi))$ and $\Phi(r_M(\psi))$ for several values of m and q , where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal. For each combination of m and q , ψ and s are such that $\Phi(r_C(\psi)) = 0.05$ and $\hat{\psi}_C = 1$. The numerical results confirm the theoretical ones. As an indication, although arbitrary, for each value of q , we put in bold face the cell corresponding to the smallest value of m that gives a probability within 0.01 of the target value. The modified profile likelihood gives practically the same results as the conditional, never requiring m larger than 6. On the contrary, the profile likelihood needs very large values of m , even for moderately large values of q .

Table 1 about here

Example 3: Inference about common odds ratio in 2×2 tables. Let us consider q independent pairs of independent binomial variables (Y_{i1}, Y_{i2}) , with $Y_{i1} \sim Bi(1, p_{i1})$ and $Y_{i2} \sim Bi(m, p_{i2})$. Considering the parameterization $\lambda_i = \log\{p_{i2}/(1 - p_{i2})\}$ and $\psi = \log\{p_{i1}/(1 - p_{i1})\} - \log\{p_{i2}/(1 - p_{i2})\}$, the model is again a full rank exponential family, as in Example 1, with components of the sufficient statistic $u = \sum_i^q y_{i1}$ and $v_i = y_{i1} + y_{i2}$, $i = 1, \dots, q$. This model may arise in case-control studies, in which we have 1 case and m controls in each table, and where interest is on studying the influence of some risk factor. The conditional likelihood is a non-central hypergeometric distribution (see for instance Example 6.1 in Davison, 1988). Profile and modified profile likelihood are easily computed using standard software for generalized linear models. Pierce & Peters (1992) studied in detail higher-order methods in this setting. Here, we propose only a particular example that is intended to show the role of m and q on the accuracy of the approximation to the conditional likelihood given by the profile and modified profile likelihoods, rather than being a detailed exploration. In fact, Table 2 reports the probabilities

$\Phi(r_P(\psi_l)), 1 - \Phi(r_P(\psi_u))$ and $\Phi(r_M(\psi_l)), 1 - \Phi(r_M(\psi_u))$, where ψ_l and ψ_u are such that $\Phi(r_C(\psi_u)) = 1 - \Phi(r_C(\psi_l)) = 0.05$. In order to have comparable settings, we considered only odd values of m and tables with $v_i = (m + 1)/2$, i.e. tables with the same number of successes and failures. Moreover, we assumed that u is equal to 1.645 standard deviations below its expected value, computed when $\psi = 0$, that is $u = q/2 - 1.645 \sqrt{q/4}$.

Table 2 about here

The results show a very accurate behavior of $r_M(\psi)$, even for large values of q and moderate values of m . This is not true for what concerns the profile likelihood, even though its accuracy does not degenerate for large values of q , as seen in Table 1. As noted also in Pierce & Peters (1992), the accuracy of the approximations may depend also on the observed value of u . Indeed, u steps by 1 from 0 to q . In the same setting of Table 2, the behavior of the three likelihoods as functions of u is symmetric around $q/2$. Figure 1 shows conditional, profile and modified profile relative loglikelihoods in the case $q = 100$, $m = 5$ and with values of u equal to 5, 30 and 45. The accuracy of the modified profile likelihood tends to be slightly worse when u approaches its boundary, while it is otherwise a really accurate approximation of the conditional likelihood. On the contrary, the profile likelihood gives reasonable results only when u is close to its mid range.

Figure 1 about here

Example 4: Matched gamma pairs. Let Y_{ij1}, Y_{ij2} be independent exponential random variables with mean ψ/λ_i and $\psi\lambda_i$, respectively. This is equivalent to considering $Y_{i1} = \sum_{j=1}^m Y_{ij1}$ and $Y_{i2} = \sum_{j=1}^m Y_{ij2}$ as matched gamma pairs with shape m and scale, respectively, λ_i/ψ and $1/(\psi\lambda_i)$. In this case, there is no exact conditional or marginal

likelihood for ψ . Hence, we compare the profile likelihood and a modification of it, through a simulation study. Since the parameters are orthogonal, we use the approximate conditional likelihood of Cox & Reid (1987). This coincides with the modified profile likelihood, since $\hat{\lambda}_\psi = \hat{\lambda}$. The profile and modified profile loglikelihoods are

$$l_P(\psi) = -2mq \log \psi - \frac{2}{\psi} \sum_{i=1}^q (y_{i1}y_{i2})^{1/2}, \quad l_M(\psi) = l_P(\psi) + \frac{1}{2} q \log \psi,$$

and the maximum likelihood estimates are $\hat{\psi} = (mq)^{-1} \sum_i (y_{i1}y_{i2})^{1/2}$ and $\hat{\psi}_M = \frac{4m}{4m-1} \hat{\psi}$. Cox & Reid (1992) show that $\hat{\psi}_M$ is less biased than $\hat{\psi}$. Here, we compare the empirical distribution of $r_P(\psi)$ and $r_M(\psi)$, with the standard normal distribution, in simulations with various values of m and q . We note that, in this case, the actual cluster sample size is $2m$. As an example, Table 3 reports coverage probabilities of the 0.05 quantile of the standard normal distribution for r_P and r_M . Similar results were found on the other tail. The parameter of interest is $\psi = 1$. The nuisance parameters were randomly chosen as integers between 1 and 10. Note that the accuracy of $r_M(\psi)$ is not very affected by large numbers of nuisance parameters, while the accuracy of $r_P(\psi)$ tends to degenerate for moderately large values of q .

Table 3 about here

Example 5: Loblolly data. Let us consider a data set concerning the growth of Loblolly pine trees (see Appendix A.13 of Pinheiro & Bates, 2000). The data consist of $q = 14$ trees with different seed source and for each of them we have $m = 6$ observations for the height (in ft), y , with respect to 6 different ages (in years) of the tree, $x = (3, 5, 10, 15, 20, 25)$. We assume $Y_{ij} = \mu(\beta_i; x_j) + \sigma \varepsilon_{ij}$, where ε_{ij} are independent standard normal random variables and

$$\mu(\beta_i; x_j) = \beta_{1i} + (\beta_{2i} - \beta_{1i}) \exp\{-e^{\beta_{3i}} x_j\}.$$

Here we consider $\psi = \sigma^2$ and $\lambda_i = \beta_i$. The maximum likelihood estimates $\hat{\beta}_i$ of the β_i 's are obtained using nonlinear least squares. The maximum likelihood estimate of σ^2 is $\hat{\sigma}^2 = (mq)^{-1} \sum_{i=1}^q \sum_{j=1}^m (y_{ij} - \mu(\hat{\beta}_i; x_j))^2$ and the profile loglikelihood is $l_P(\sigma^2) = -\frac{mq}{2} \{\log \sigma^2 + \hat{\sigma}^2/\sigma^2\}$. Since $\hat{\beta}_i = \hat{\beta}_{i\sigma^2}$, the modified profile loglikelihood coincides with the Cox & Reid adjusted profile loglikelihood and is $l_M(\sigma^2) = l_P(\sigma^2) + \frac{3}{2} q \log \sigma^2$. This gives the estimate $\hat{\sigma}_M^2 = m\hat{\sigma}^2/(m-3) = 2\hat{\sigma}^2$. In this case, we have $\hat{\sigma}^2 = 0.2453$ and $\hat{\sigma}_M^2 = 0.4906$. Figure 2 shows the relative loglikelihoods for σ^2 together with the 0.95 confidence intervals based on the asymptotic distribution of the likelihood ratio statistics.

Figure 2 about here

A simulation with 10,000 replications, $\beta_i = \hat{\beta}_i$ and $\sigma^2 = 0.5$, gave coverage probabilities for the nominal 0.95 confidence interval equal to 0.025 for the profile likelihood and 0.951 for the modified profile likelihood.

Suppose now that ε_{ij} are independent random variables with Student distribution t_ν with $\nu = 3$ degrees of freedom. In this case, $\hat{\beta}_{i\sigma^2}$ is no longer equal to $\hat{\beta}_i$, even though β_i is still orthogonal to σ^2 . The modified profile likelihood may be computed using the approximation of sample space derivatives given in § 9.5.3 of Severini (2000), which is due to Fraser & Reid (1995). In particular, using $(y_{ij} - \mu(\beta_i; x_j))/\sigma$ as pivotal quantity, we have that the approximation $\tilde{l}_{\beta_i; \hat{\beta}_i}(\sigma^2, \beta_i)$, which is a three by three matrix, is proportional to

$$\sum_{j=1}^m \frac{\nu\sigma^2 - (y_{ij} - \mu(\beta_i; x_j))^2}{\{\nu\sigma^2 + (y_{ij} - \mu(\beta_i; x_j))^2\}^2} \mu'(\beta_i; x_j) \mu'(\hat{\beta}_i; x_j)^T,$$

where $\mu'(\beta_i; x_j) = \partial\mu(\beta_i; x_j)/\partial\beta_i$ is a three by one vector. In this case, the estimates are $\hat{\sigma}^2 = 0.0833$ and $\hat{\sigma}_M^2 = 0.3396$. Figure 3 shows the relative loglikelihoods. Note that the modified profile likelihood

almost coincides with the Cox & Reid adjusted profile likelihood, but the latter is not invariant with respect to interest respecting reparameterizations.

Figure 3 about here

6. DISCUSSION

The main message of this paper is that, in models in which the number of nuisance parameters is substantial relative to the sample size, the profile likelihood fails, while modified profile likelihoods still perform reasonably. Conditions are stated on the asymptotic rate of the number of nuisance parameters and the cluster sample size in order to have the usual asymptotic results. Indeed, the profile likelihood requires $1/m = o(q^{-1})$, while the corresponding condition for modified profile likelihoods is $1/m = o(q^{-1/3})$. Three further comments follow.

- (a) Modified profile likelihoods typically have a location and spread adjustment on the profile likelihood; see for instance Figures 2 and 3. In this paper we concentrated mainly on the location adjustment. However, even when this is not necessary, we can have a sensible spread adjustment, since usually j_P overestimates the partial information. As an example, consider Y_{ij} as independent normal with mean μ and variance σ_i^2 (see Example 9.18 in Severini, 2000). When $m_i = m$ we have $\hat{\mu} = \hat{\mu}_M$, but the confidence intervals based on W_P tend to be too narrow. For instance, in a simulation with 10.000 replications, $m = 5$ and $q = 25$, we had an actual coverage of 0.866 for W_P and of 0.947 for W_M , when the nominal level was 0.95.
- (b) When ψ is a scalar, we could also use the modified directed likelihood (Barndorff-Nielsen, 1986; Fraser *et al.*, 1999) to compute p -values and confidence intervals. The modified directed likeli-

hood usually gives very accurate results even in rather extreme situations. This is probably due to its large deviation properties. Pierce & Peters (1992) show that the modified directed likelihood can be written as $r_P + NP + INF$, where NP is a nuisance parameter adjustment and INF is an information adjustment. In $m \times q$ -asymptotics, using the results in this paper together with some other simple expansions, it is possible to see that $NP = O_p(\sqrt{q/m})$, while INF has the usual order $O_p(n^{-1/2})$. This is true regardless of the nature of the sequence $\{q, m\}$. Moreover, it can be shown that $r_M = r_P + NP + o_p(1)$, provided that $1/m = o(q^{-1/3})$. This means that, under this condition, r_M tends to be very close to the modified directed likelihood. See also Sartori *et al.* (1999) for some simulation results.

- (c) Under the same conditions defined in Severini (2002), the results of Section 4 on $\hat{\psi}$ and $\hat{\psi}_M$ extend to profile and modified profile estimating functions.

APPENDIX

Some technical details

Consider the central limit theorem for independent variables, where X_i are independent random variables with mean μ_i and variance σ_i^2 , $i = 1, \dots, q$. Then

$$T_q = \frac{\sum_{i=1}^q (X_i - \mu_i)}{C_q} \xrightarrow{p} N(0, 1)$$

where $C_q = (\sum_{i=1}^q \sigma_i^2)^{1/2}$. This implies that $T_q = O_p(1)$ and

$$\sum_{i=1}^q X_i = O\left(\sum_{i=1}^q \mu_i\right) + O_p(C_q). \quad (11)$$

Let us use this result to evaluate the order of terms in (6). We have already seen that $U_{\psi|\lambda}$ is $O_p(n^{1/2})$. The term B_P is of order $O_p(q)$ since $E_\theta(B_P^i) = O(1)$ and $V_\theta(B_P^i) = O(1)$. This implies that in (11) we have $\sum_{i=1}^q \mu_i = O(q)$ and $C_q = O(q^{1/2})$, giving $B_P = O(q) + O_p(q^{1/2}) = O_p(q)$. The term R_P is such that $E_\theta(R_P^i) = O(m^{-1})$ and $V_\theta(R_P^i) = O(m^{-1})$. This implies that $R_P = O_p(q/m) + O_p\{(q/m)^{1/2}\} = O_p[\max\{q/m, (q/m)^{1/2}\}]$.

The modified profile score function can be written in the form

$$U_M = U_{\psi|\lambda} + B_M + R_P,$$

where $B_M = B_P + \partial M / \partial \psi = \sum_{i=1}^q (B_P^i + \partial M^i / \partial \psi) = \sum_{i=1}^q B_M^i$. This follows from (6) and (4). Using (5), it is straightforward to see that $E_\theta(B_M^i) = O(m^{-1})$ and, using the delta method, that $V_\theta(B_M^i) = O(m^{-1})$. This means that B_M and R_P are quantities of the same order. Hence, we can write $U_M = U_{\psi|\lambda} + R_M$, where $R_M = B_M + R_P = O_p[\max\{q/m, (q/m)^{1/2}\}]$.

Finally, let us consider the relation between the partial information and the observed profile information $j_P(\psi) = \sum_{i=1}^q j_P^i(\psi)$, where

$$j_P^i(\psi) = j_{\psi\psi}^i(\psi, \hat{\lambda}_{i\psi}) - j_{\lambda_i\lambda_i}^{-1}(\psi, \hat{\lambda}_{i\psi}) j_{\psi\lambda_i}^2(\psi, \hat{\lambda}_{i\psi})$$

is the observed profile information in the i -th cluster. Expanding each single term around (ψ, λ) , using an expansion for $(\hat{\lambda}_{i\psi} - \lambda_i)$ (see, for instance formula (9.87) in Pace & Salvan, 1997), we have that

$$j_P(\psi) = i_{\psi\psi|\lambda} + O_p(q) + O_p(n^{1/2}). \quad (12)$$

Moreover, since $i_{\psi\psi|\lambda} = n \bar{i}_{\psi\psi|\lambda}$, we can write

$$j_P(\psi) = i_{\psi\psi|\lambda} \{1 + O_p(m^{-1}) + O_p(n^{-1/2})\} = i_{\psi\psi|\lambda} \{1 + o_p(1)\}. \quad (13)$$

The relation between the observed information and the partial information is the same even for a modified profile likelihood. In fact, since

derivatives of $M(\psi)$ are of order $O_p(q)$, we have

$$\begin{aligned} j_M(\psi) &= j_P(\psi) - \partial^2 M(\psi) / \partial \psi^2 = j_P(\psi) + O_p(q) \\ &= i_{\psi\psi|\lambda} + O_p(q) + O_p(n^{1/2}), \end{aligned}$$

where the last step follows from (12).

REFERENCES

- ANDERSEN, E.B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *J. R. Statist. Soc. B* **32**, 283-301.
- ANDERSEN, E.B. (1971). Asymptotic properties of conditional likelihood ratio tests. *J. Amer. Statist. Assoc.*, **66**, 630-633.
- BARNDORFF-NIELSEN, O.E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343-65.
- BARNDORFF-NIELSEN, O.E. (1986). Inference on full and partial parameters based on standardized signed log likelihood ratio. *Biometrika* **73**, 307-22.
- 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. (1995). Stable and invariant adjusted profile likelihood and directed likelihood for curved exponential models. *Biometrika* **82**, 489-500.
- BARNDORFF-NIELSEN, O.E. (1996). Two index asymptotics. In *Frontiers in Pure and Applied Probability II: Proceedings of the Fourth Russian-Finnish Symposium Prob. Th. Math. Statist.*, Ed. A. Melnikov, pp 9-20. Moscow: TVP Science.
- BARNDORFF-NIELSEN, O.E. & COX, D.R. (1994). *Inference and Asymptotics*. London: Chapman and Hall.
- 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. (1992). A note on the difference between profile and modified profile likelihood. *Biometrika*, **79**, 408-11.
- DAVISON, A.C. (1988). Approximate conditional inference in generalized linear models. *J. R. Statist. Soc. B* **50**, 445-61.
- DAVISON, A.C. (1992). Discussion of the paper by D.A. Pierce & D. Peters (1992). *J. R. Statist. Soc. B* **54**, 726-7.
- DiCICCIO, T.J. & STERN, S. (1994). Constructing approximately standard normal pivots from signed roots of adjusted likelihood ratio statistics. *Scand. J. Statist.* **21**, 447-60.
- 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.
- FRASER, D. A. S. & REID, N. (1995). Ancillaries and third-order significance. *Utilitas Mathematica*, **47**, 33-53.
- MCCULLAGH, P. & TIBSHIRANI, R. (1990). A simple method for the adjustment of profile likelihoods. *J. R. Statist. Soc. B*, **52**, 325-44.
- NEYMAN, J. & SCOTT, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, **16**, 1-32.
- PACE, L. & SALVAN, A. (1997). *Principles of statistical inference from a neo-Fisherian perspective*. World Scientific Publishing Co., Singapore.
- PIERCE, D.A. & PETERS, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families (with Discussion). *J. R. Statist. Soc. B* **54**, 701-37.
- PINHEIRO, J.C. & BATES, D.M. (2000). *Mixed-Effects models in S and S-PLUS*. New York: Springer-Verlag.
- PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.*, **16**, 356-66.
- 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.A. (1998). An approximation to the modified profile likelihood function. *Biometrika* **85**, 403-11.

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

SEVERINI, T.A. (2002). Modified estimating functions. *Biometrika*, to appear.

Table 1: Common shape gamma samples $\phi(x; \psi)$ and $\phi(x; \psi)$ with ψ such that $\phi(x; \psi) = 0.05$ in samples with $n = 1$. For each ψ values in bold face correspond to the smallest ψ for ψ and ψ such that the probability is within 0.01 of 0.05.

m		q				
		4	8	16	64	128
3	r_P	0.190	0.299	0.487	0.952	0.999
	r_M	0.053	0.055	0.058	0.070	0.080
4	r_P	0.159	0.239	0.383	0.866	0.988
	r_M	0.052	0.053	0.055	0.062	0.068
5	r_P	0.141	0.206	0.322	0.777	0.962
	r_M	0.052	0.052	0.054	0.058	0.062
6	r_P	0.129	0.184	0.281	0.698	0.924
	r_M	0.051	0.052	0.053	0.056	0.059
7	r_P	0.121	0.168	0.252	0.632	0.879
	r_M	0.051	0.051	0.052	0.055	0.057
10	r_P	0.105	0.140	0.201	0.494	0.748
	r_M	0.051	0.051	0.051	0.053	0.054
50	r_P	0.071	0.081	0.098	0.174	0.260
	r_M	0.050	0.050	0.050	0.050	0.050
100	r_P	0.064	0.070	0.081	0.126	0.174
	r_M	0.050	0.050	0.050	0.050	0.050
200	r_P	0.060	0.064	0.071	0.098	0.126
	r_M	0.050	0.050	0.050	0.050	0.050
1000	r_P	0.054	0.056	0.059	0.069	0.077
	r_M	0.050	0.050	0.050	0.050	0.050

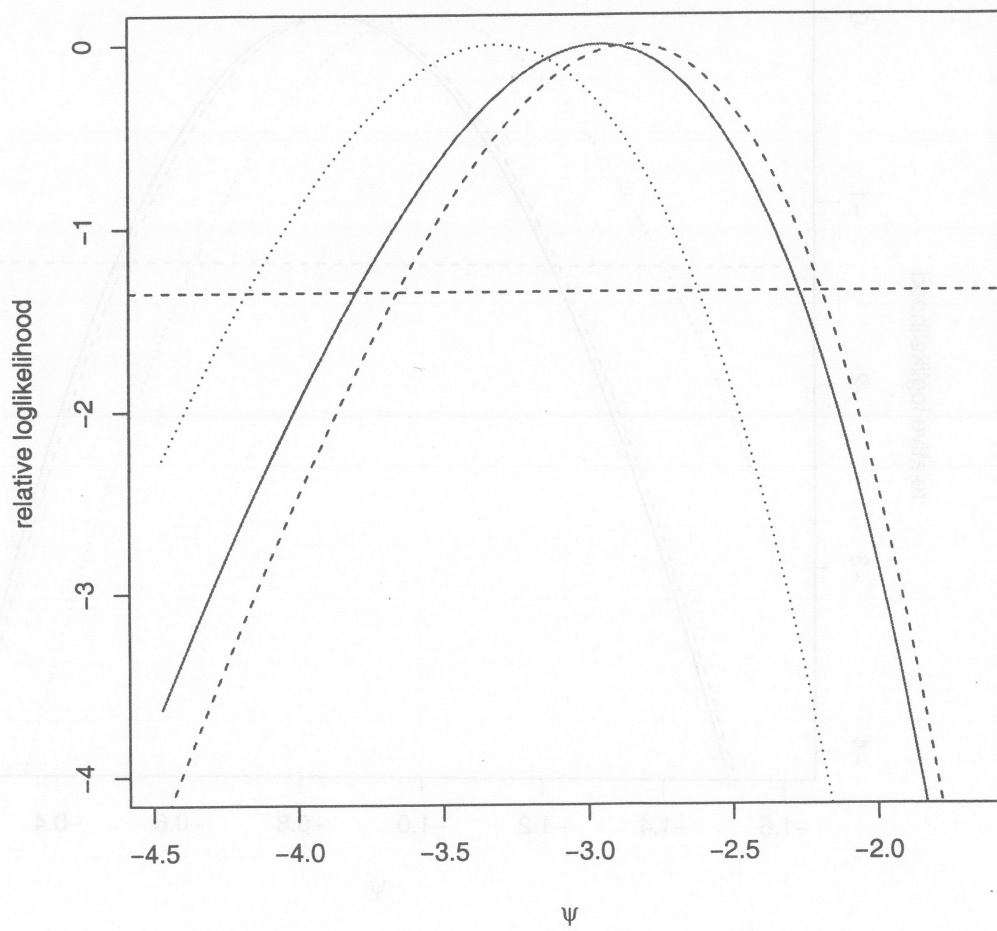
Table 1: Common shape gamma samples. $\Phi(r_P(\psi))$ and $\Phi(r_M(\psi))$ with ψ such that $\Phi(r_C(\psi)) = 0.05$ in samples with $\hat{\psi}_C = 1$. For each q , values in bold face correspond to the smallest m , for r_P and r_M , such that the probability is within 0.01 of 0.05.

m		q				
		20	50	100	300	600
1	r_P	0.494,0.007	0.499,0.009	0.480,0.011	0.490,0.011	0.495,0.010
	r_M	0.130,0.027	0.157,0.028	0.160,0.030	0.167,0.029	0.169,0.029
3	r_P	0.150,0.026	0.164,0.028	0.163,0.030	0.167,0.029	0.170,0.029
	r_M	0.055,0.045	0.063,0.045	0.066,0.045	0.067,0.045	0.068,0.045
5	r_P	0.102,0.034	0.110,0.035	0.110,0.036	0.113,0.036	0.114,0.036
	r_M	0.051,0.048	0.055,0.048	0.056,0.048	0.057,0.048	0.057,0.048
7	r_P	0.085,0.038	0.090,0.039	0.091,0.040	0.092,0.040	0.093,0.039
	r_M	0.051,0.049	0.053,0.049	0.053,0.049	0.054,0.049	0.054,0.049
9	r_P	0.076,0.040	0.080,0.041	0.080,0.042	0.082,0.042	0.082,0.042
	r_M	0.050,0.049	0.052,0.049	0.052,0.049	0.052,0.049	0.053,0.049

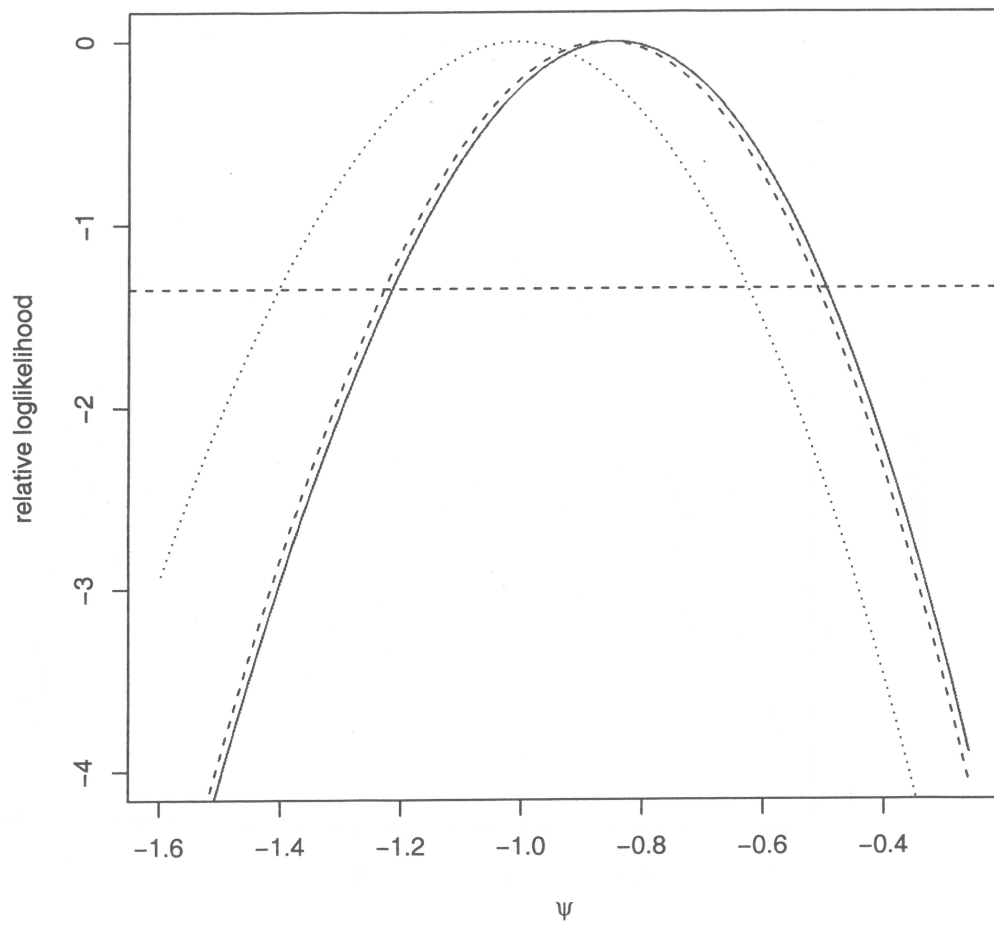
Table 2: Case-control study. One-sided non-coverage probabilities of the 0.90 conditional confidence interval. We consider q tables with 1 case and m controls and with $v_i = (m + 1)/2$ in each table. We fix u equal to $q/2 - 1.645\sqrt{q/4}$. For each q , values in bold face correspond to the smallest m , for r_P and r_M , such that the probabilities are both within 0.01 of 0.05.

m		q				
		4	8	16	64	128
1	r_P	0.210	0.278	0.408	0.836	0.980
	r_M	0.049	0.039	0.034	0.017	0.009
2	r_P	0.148	0.186	0.270	0.625	0.861
	r_M	0.054	0.049	0.048	0.033	0.032
3	r_P	0.123	0.159	0.206	0.494	0.745
	r_M	0.054	0.053	0.048	0.046	0.044
5	r_P	0.102	0.123	0.164	0.349	0.551
	r_M	0.054	0.053	0.053	0.041	0.046
10	r_P	0.083	0.098	0.117	0.237	0.364
	r_M	0.053	0.055	0.048	0.053	0.051
50	r_P	0.062	0.068	0.077	0.107	0.143
	r_M	0.051	0.052	0.053	0.049	0.053

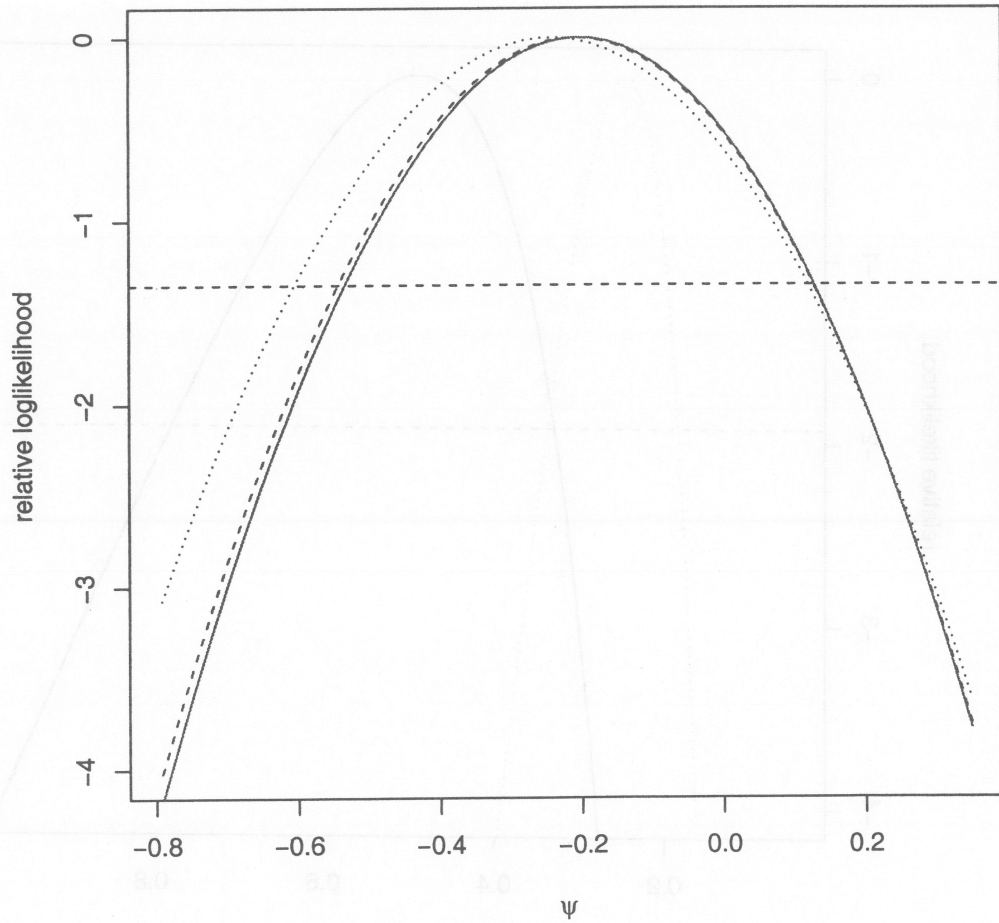
Table 3: Matched Gamma pairs. Coverage probabilities of the 0.05 quantile of the standard normal distribution for $r_P(\psi)$ and $r_M(\psi)$ in simulations with 100.000 replications each, and with various values of m and q . The parameter of interest is $\psi = 1$. For each q , values in bold face are corresponding to the smallest m , for r_P and r_M , such that the probability is within 0.01 of 0.05.



(a)



(b)



(c)

Figure 1: Case control study. Relative loglikelihoods for ψ : profile (dotted), modified profile (dashed) and conditional (solid). The horizontal dashed line gives the 0.90 confidence interval. We consider $q = 100$, $m = 5$, $v_i = 3$, $i = 1, \dots, q$ and three values of u : 5 (a), 30 (b) and 45 (c).

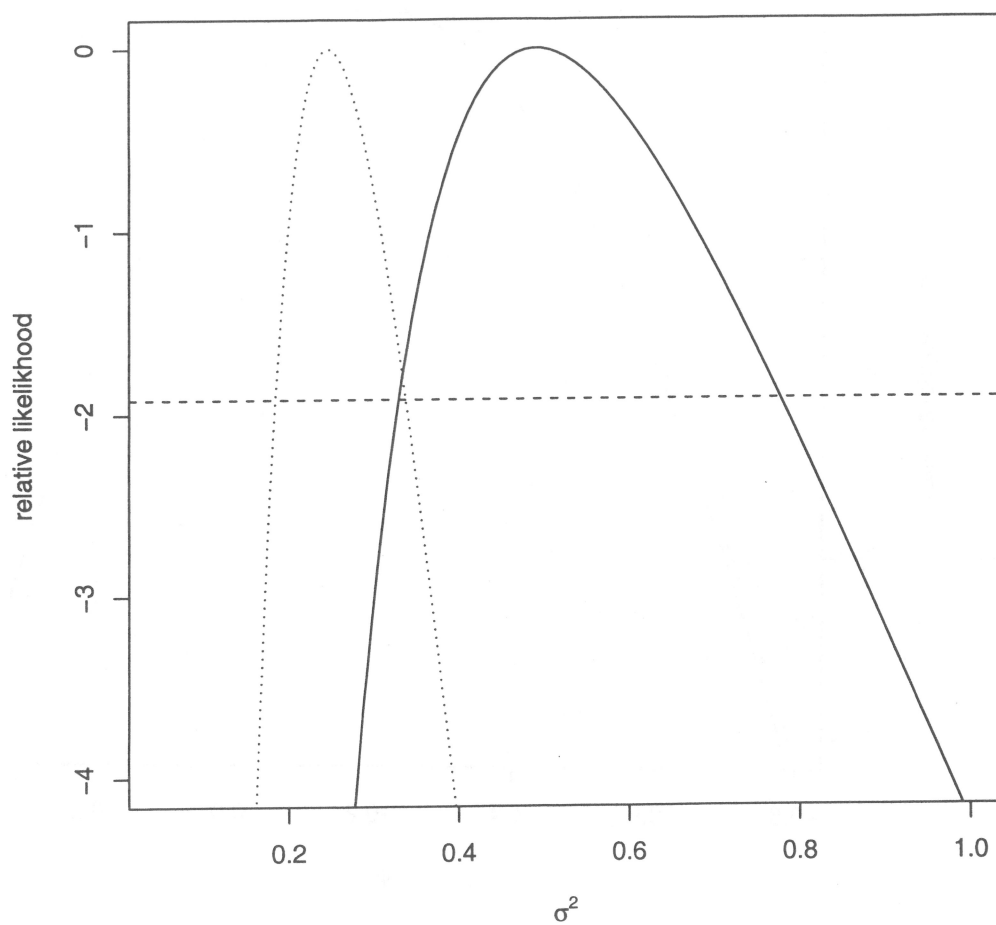


Figure 2: Loblolly data. Relative loglikelihoods for σ^2 when the distribution is normal: profile (dotted), modified profile (solid). The horizontal dashed line gives the 0.95 confidence interval.

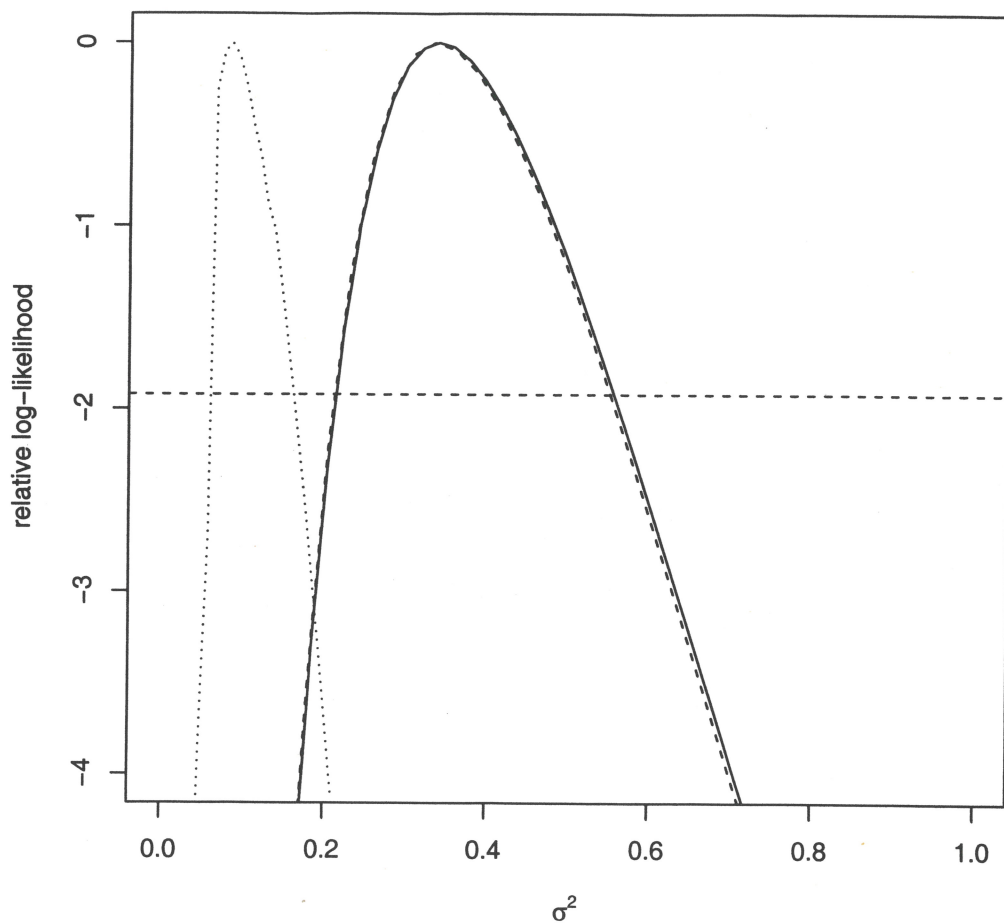


Figure 3: Loblolly data. Relative loglikelihoods for σ^2 when the distribution is t_3 : profile (dotted), modified profile (solid), Cox & Reid (dashed). The horizontal dashed line gives the 0.95 confidence interval.

