J. R. Statist. Soc. B (1996)
58, *No.* 1, *pp.* 189–203

Information Bias and Adjusted Profile Likelihoods

By THOMAS J. DiCICCIO†,
Cornell University, Ithaca, USA

MICHAEL A. MARTIN,
Australian National University, Canberra, Australia, and Stanford University, USA

STEVEN E. STERN
Australian National University, Canberra, Australia

and

G. ALASTAIR YOUNG
University of Cambridge, UK

[Received June 1993. Final revision December 1994]

SUMMARY

The bias and information bias of the ordinary profile score statistic are both typically of order $O(1)$. Several additive adjustments to the profile score statistic that reduce its bias to order $O(n^{-1})$ have been proposed. In certain situations, the information bias of these adjusted profile score statistics is also reduced to order $O(n^{-1})$; specifically, the modified profile likelihood of Barndorff-Nielsen yields an adjusted profile score statistic having both reduced bias and reduced information bias. In general, however, a bias reducing adjustment to the profile score statistic will not automatically reduce the order of the information bias as well. In this paper, an analytical formula is obtained for the information bias of a general bias-adjusted profile score statistic. This formula is used to compare bias reducing adjustments to the profile score statistic, as well as to construct further adjustments that reduce the information bias to $O(n^{-1})$. Several examples are presented to illustrate use of the formula for information bias. In particular, the information bias formula may be utilized in a criterion for choosing between orthogonal parameterizations for the conditional profile likelihood of Cox and Reid.

Keywords: ADDITIVE ADJUSTMENT; ASYMPTOTIC EXPANSION; BIAS REDUCTION; EXPONENTIAL FAMILY; LOCATION–SCALE FAMILY; NUISANCE PARAMETER; ORTHOGONAL PARAMETER; PARAMETERIZATION INVARIANCE; PROFILE SCORE STATISTIC; SCALING ADJUSTMENT

1. INTRODUCTION

Consider an observed random variable $Y = (Y_1, \ldots, Y_n)$ having probability distribution that depends on an unknown parameter $\theta = (\theta^1, \ldots, \theta^{p+q})$, and let $L(\theta)$ denote the log-likelihood function for θ based on Y. Suppose that θ is partitioned as $\theta = (\psi, \phi)$, where ψ is the p-dimensional parameter of interest and ϕ is a nuisance parameter. Let $\hat{\theta} = (\hat{\psi}, \hat{\phi})$ be the overall maximum likelihood estimator of θ, and let $\tilde{\theta}(\psi) = \{\psi, \tilde{\phi}(\psi)\}$ be the constrained maximum likelihood estimator of θ for a given ψ. Inference for ψ is often based on the profile log-likelihood function $M(\psi) = L\{\tilde{\theta}(\psi)\}$. For example, $M(\psi)$ is maximized at $\hat{\psi}$, and the profile likelihood ratio statistic $W(\psi) = 2\{M(\hat{\psi}) - M(\psi)\}$ is typically distributed as χ_p^2 up to an error of order $O(n^{-1})$. Generally, the error in the χ^2-approximation to the distribution of W can be reduced to order $O(n^{-2})$ by Bartlett correction.

†*Address for correspondence*: Department of Social Statistics, Ives Hall, Cornell University, Ithaca, NY 14853, USA.

0035–9246/96/58189

Since the construction of $M(\psi)$ involves estimation of the nuisance parameters, the profile log-likelihood function does not behave exactly like a true log-likelihood function. In particular, the profile score function $\partial M(\psi)/\partial\psi$ has bias

$$E_\theta\left\{\frac{\partial M(\psi)}{\partial\psi}\right\}$$

and information bias

$$E_\theta\left\{\frac{\partial M(\psi)}{\partial\psi^{\mathrm{T}}}\,\frac{\partial M(\psi)}{\partial\psi}\right\}+E_\theta\left[\frac{\partial}{\partial\psi^{\mathrm{T}}}\left\{\frac{\partial M(\psi)}{\partial\psi}\right\}\right],$$

both of which do not typically vanish, but rather they are of order $O(1)$. Godambe (1960) and Lindsay (1982) have discussed information unbiasedness.

Several researchers have proposed additive adjustments to the profile score function that reduce its bias to order $O(n^{-1})$, including Bartlett (1955), Barndorff-Nielsen and Cox (1979), Barndorff-Nielsen (1983, 1994), Cox and Reid (1987, 1993), McCullagh and Tibshirani (1990), Barndorff-Nielsen and Chamberlin (1994) and DiCiccio and Stern (1993). The bias reducing properties of these adjustments to the profile score function are discussed further by Liang (1987), Levin and Kong (1990), Ferguson *et al.* (1991) and Cox and Reid (1992).

The additive adjustments are often obtained by replacing the profile log-likelihood function $M(\psi)$ with a function of the form

$$\bar{M}(\psi)=M(\psi)+B(\psi), \tag{1}$$

where $B(\psi)$ is a suitably smooth function having derivatives of order $O_p(1)$. The resultant adjustment to the profile score function is $\partial B(\psi)/\partial\psi$. The estimate $\bar{\psi}$ of ψ obtained by maximizing the adjusted profile log-likelihood function $\bar{M}(\psi)$ satisfies $\bar{\psi}=\hat{\psi}+O_p(n^{-1})$, and the χ^2_p-approximation to the distribution of the adjusted profile likelihood ratio statistic $\bar{W}(\psi)=2\{\bar{M}(\bar{\psi})-\bar{M}(\psi)\}$ has error of order $O(n^{-1})$. Under mild assumptions on the adjustment function $B(\psi)$, DiCiccio and Stern (1994) showed that $\bar{W}(\psi)$ is Bartlett correctable, and they derived a formula for the Bartlett adjustment factor.

Barndorff-Nielsen (1983) proposed an additive adjustment that involves an exact or approximate ancillary in cases where $\hat{\theta}$ is not minimal sufficient. The adjustment proposed by Cox and Reid (1987) was developed under the assumption that the parameters of interest are orthogonal to the nuisance parameters. McCullagh and Tibshirani (1990) have discussed the use of simulation to construct both additive and scaling adjustments for the profile score function that are designed to reduce both the bias and the information bias. They provided an analytical expression for an additive adjustment that reduces the bias to order $O(n^{-1})$, and, in the context of exponential families, they derived an analytical expression for a scaling adjustment that reduces information bias to the order $O(n^{-1})$.

This paper concerns the information bias that arises from the use of bias reducing additive adjustments to the profile score function in general parametric models. We give an analytical expression for the information bias that results when such adjustments are applied. Although the bias is reduced to order $O(n^{-1})$ by these additive adjustments, the information bias generally remains of order $O(1)$. However, the analytical expression for the information bias allows a construction of further adjustments that reduce the information bias to order $O(n^{-1})$. Furthermore, this expression allows investigation of the effect that previously proposed bias reducing additive adjustments have on information bias for specific models.

The analytical expression for information bias is derived in Section 2, and it is expressed in terms of expectations of the derivatives of the log-likelihood function. Further adjustments to the profile score function that reduce information bias are also developed in Section 2. In Section 3, the additive adjustments proposed by Cox and Reid (1987), Barndorff-Nielsen (1983, 1994) and DiCiccio and Stern (1993) are examined and compared with regard to information bias. Specific examples that facilitate comparison of these various bias reducing adjustments are given in Section 4. In particular, some cases of practical interest are presented for which these adjustments simultaneously reduce both bias and information bias to order $O(n^{-1})$. Section 5 concerns the use of information bias as the basis of a criterion to choose between the possible orthogonal parameterizations in a given model. This criterion is closely related to a criterion proposed by Cox and Reid (1989).

2. CALCULATION OF INFORMATION BIAS

Some necessary notation is summarized at the outset. In the formulae that follow, the standard conventions for denoting arrays and sums are employed. In using these conventions, it is to be understood that the indices $a, b, c, \ldots$ range over $1, \ldots, p$, that the indices $i, j, k, \ldots$ range over $p+1, \ldots, p+q$, and that the indices $r, s, t, \ldots$ range over $1, \ldots, p+q$. Differentiation is denoted by subscripts, so that $L_r(\theta) = \partial L(\theta)/\partial\theta^r$, $L_{rs}(\theta) = \partial^2 L(\theta)/\partial\theta^r\partial\theta^s$, $M_a(\psi) = \partial M(\psi)/\partial\psi^a$, $M_{ab}(\psi) = \partial^2 M(\psi)/\partial\psi^a\partial\psi^b$, etc. Define $\lambda_{rs} = E_\theta\{L_{rs}(\theta)\}$, $\lambda_{rst} = E_\theta\{L_{rst}(\theta)\}$, $\lambda_{rs/t} = \partial\lambda_{rs}/\partial\theta^t$, etc., and assume that these quantities are of order $O(n)$. Further, the zero-mean variables $l_r = L_r(\theta)$, $l_{rs} = L_{rs}(\theta) - \lambda_{rs}$, etc. are assumed to be of order $O_p(n^{1/2})$. These assumptions are typically satisfied in practice. Let (λ^{rs}) be the $(p+q) \times (p+q)$ matrix inverse of (λ_{rs}), and let (σ_{ab}) be the matrix inverse of (λ^{ab}), the upper left-hand $p \times p$ submatrix of (λ^{rs}). Set $\nu^{rs} = \lambda^{rs} - \lambda^{ra}\lambda^{sb}\sigma_{ab}$. The entries of the matrix (ν^{rs}) are all 0, except for the lower right-hand $q \times q$ submatrix (ν^{ij}), which is the matrix inverse of (λ_{ij}).

The bias of the profile score function is usually taken into account by making an additive adjustment $B_a(\psi)$ to $M_a(\psi)$, which yields

$$\bar{M}_a(\psi) = M_a(\psi) + B_a(\psi) \qquad (a = 1, \ldots, p), \tag{2}$$

where the quantities $B_a(\psi)$ $(a = 1, \ldots, p)$ are of order $O_p(1)$. Define the expectations $\beta_a = E_\theta\{B_a(\psi)\}$, $\beta_{ab} = E_\theta\{B_{a/b}(\psi)\}$, etc., and put $b_a = B_a(\psi) - \beta_a$, $b_{ab} = B_{a/b}(\psi) - \beta_{ab}$, etc., where $B_{a/b}(\psi) = \partial B_a(\psi)/\partial\psi^b$. It is assumed that the means $\beta_a, \beta_{ab}, \ldots$ are of order $O(1)$ and that the joint cumulants of $nb_a, nb_{ab}, l_r, l_{rs}, \ldots$ are of order $O(n)$; specifically, $b_a, b_{ab}, \ldots$ are assumed to be of order $O_p(n^{-1/2})$. These assumptions are also typically satisfied in practice. In the case $p > 1$, the derivatives $B_{a/b}(\psi)$ and $B_{b/a}(\psi)$ do not necessarily coincide for $a \neq b$, and indeed, the existence of a function $B(\psi)$ for which $B_a(\psi) = \partial B(\psi)/\partial\psi^a$ $(a = 1, \ldots, p)$ is not guaranteed. Consequently, for $a \neq b$, β_{ab} and b_{ab} differ from β_{ba} and b_{ba} in general; moreover, there need not exist a function $\bar{M}(\psi)$ such that $\bar{M}_a(\psi) = \partial\bar{M}(\psi)/\partial\psi^a$ $(a = 1, \ldots, p)$.

McCullagh and Tibshirani (1990) showed that

$$E_\theta\{M_a(\psi)\} = -\rho_a(\theta) + O(n^{-1}),$$

where

$$\rho_a = \sigma_{ab}\,\lambda^{br}\,\nu^{st}(\lambda_{rs/t} - \tfrac{1}{2}\,\lambda_{rst}) = \sigma_{ab}\,\lambda^{br}\,\nu^{ij}\,(\lambda_{ri/j} - \tfrac{1}{2}\,\lambda_{rij});$$

see also Bartlett (1955). Thus, if the adjustment terms in equation (2) are chosen so that

$$\beta_a = E_\theta\{B_a(\psi)\} = \rho_a + O(n^{-1}) \qquad (a = 1, \ldots, p), \tag{3}$$

then $E_\theta\{\bar{M}_a(\psi)\}$ is of order $O(n^{-1})$. McCullagh and Tibshirani (1990) considered the case $B_a(\psi) = \rho_a\{\tilde{\theta}(\psi)\}$; this choice and others are discussed in Section 3.

Now suppose that the adjustment terms $B_a(\psi)$ $(a = 1, \ldots, p)$ satisfy equation (3), and consider another adjusted profile score function given by

$$\bar{M}^\dagger_a(\psi) = \{\delta^b_a + C^b_a(\psi)\}\,\bar{M}_b(\psi) \qquad (a = 1, \ldots, p), \tag{4}$$

where δ^b_a is Kronecker's delta, $C^b_a(\psi) = \zeta^b_a(\theta) + O_p(n^{-3/2})$ and $\zeta^b_a(\theta)$ is a non-random quantity of order $O(n^{-1})$. The quantity $\bar{M}^\dagger_a(\psi)$ is a scaled version of the bias-adjusted profile score function $\bar{M}_a(\psi)$. Clearly, the bias $E_\theta\{\bar{M}^\dagger_a(\psi)\}$ is of order $O(n^{-1})$. The objective is to determine $\zeta^b_a(\theta)$ so that the information bias

$$\bar{\Delta}_{ab} = E_\theta\{\bar{M}^\dagger_a(\psi)\,\bar{M}^\dagger_b(\psi)\} + E_\theta\{\bar{M}^\dagger_{a/b}(\psi)\} \qquad (a, b = 1, \ldots, p)$$

is also of order $O(n^{-1})$. Here $\bar{M}^\dagger_{a/b}(\psi) = \partial\bar{M}^\dagger_a(\psi)/\partial\psi^b$, and it is assumed that $C^b_{a/c}(\psi) = \zeta^b_{ac}(\theta) + O_p(n^{-3/2})$, where ζ^b_{ac} is a non-random quantity of order $O(n^{-1})$.

For determining $\zeta^b_a(\theta)$, it is useful to express $\bar{\Delta}_{ab}$ in terms of the information bias of the unadjusted profile score function,

$$\Delta_{ab} = E_\theta\{M_a(\psi)\,M_b(\psi)\} + E_\theta\{M_{ab}(\psi)\}.$$

From definition (4),

$$\begin{aligned}\bar{M}^\dagger_a(\psi) &= M_a(\psi) + B_a(\psi) + \zeta^b_a M_b(\psi) + O_p(n^{-1}),\\ \bar{M}^\dagger_{a/b}(\psi) &= M_{ab}(\psi) + B_{a/b}(\psi) + \zeta^c_{ab}\,M_c(\psi) + \zeta^c_a\,M_{cb}(\psi) + O_p(n^{-1}).\end{aligned}$$

Since $M_a(\psi) = \sigma_{ab}\,\lambda^{br}\,l_r + O_p(1)$, it follows from these expansions that

$$\begin{aligned}E_\theta\{\bar{M}^\dagger_a(\psi)\,\bar{M}^\dagger_b(\psi)\} =&E_\theta\{M_a(\psi)M_b(\psi)\} + E_\theta\{M_a(\psi)\,B_b(\psi)\} + E_\theta\{B_a(\psi)\,M_b(\psi)\}\\ &+ E_\theta\{B_a(\psi)\,B_b(\psi)\} + E_\theta\{M_a(\psi)\zeta^c_b\,M_c(\psi)\} + E_\theta\{\zeta^c_a\,M_c(\psi)\,M_b(\psi)\}\\ &+ O(n^{-1}),\\ E_\theta\{\bar{M}^\dagger_{a/b}(\psi)\} =&E_\theta\{M_{ab}(\psi)\} + E_\theta\{B_{a/b}(\psi)\} + E_\theta\{\zeta^c_a\,M_{cb}(\psi)\} + O(n^{-1}).\end{aligned}$$

Standard calculations then yield

$$E_\theta\{M_a(\psi)B_b(\psi)\} = -\rho_a\rho_b + \sigma_{ac}\rho_{b/r}\lambda^{cr} - \beta_{ba} + O(n^{-1}),$$

$$E_\theta\{B_a(\psi)B_b(\psi)\} = \rho_a\rho_b + O(n^{-1}),$$

$$E_\theta\{M_a(\psi)\zeta_b^c M_c(\psi)\} = -\sigma_{ac}\zeta_b^c + O(n^{-1}),$$

$$E_\theta\{\zeta_a^c M_{cb}(\psi)\} = \zeta_a^c\sigma_{cb} + O(n^{-1}).$$

Therefore,

$$\bar{\Delta}_{ab} = \Delta_{ab} - \rho_a\rho_b + \rho_{a/r}\sigma_{bc}\lambda^{rc} + \sigma_{ac}\rho_{b/r}\lambda^{cr} - \beta_{ba} - \sigma_{ac}\zeta_b^c + O(n^{-1}). \tag{5}$$

It follows from equation (5) that the information bias $\bar{\Delta}_{ab}$ ($a, b = 1, \ldots, p$) is reduced to order $O(n^{-1})$ when

$$\begin{aligned}\zeta_a^b &= (\Delta_{ac} - \rho_a\rho_c + \rho_{a/r}\sigma_{cd}\lambda^{rd} + \sigma_{ad}\rho_{c/r}\lambda^{dr} - \beta_{ac})\lambda^{bc} + O(n^{-2}) \\ &= \Delta_{ac}\lambda^{bc} - \rho_a\rho_c\lambda^{bc} + \rho_{a/r}\lambda^{br} + \sigma_{ad}\rho_{c/r}\lambda^{bc}\lambda^{dr} - \beta_{ac}\lambda^{bc} + O(n^{-2}).\end{aligned} \tag{6}$$

Formulae for Δ_{ab} and $\rho_{a/r}$ that facilitate evaluation of equation (6) are given in Appendix A. Once ζ_a^b has been determined, there are several natural choices for $C_a^b(\psi)$, notably $C_a^b(\psi) = \zeta_a^b\{\tilde{\theta}(\psi)\}$ and $C_a^b(\psi) = \zeta_a^b(\hat{\theta})$.

For any additive adjustment that satisfies equation (3), the information bias of $\bar{M}_a(\psi)$ can be obtained from equation (5) by setting $\zeta_a^b = 0$. Thus,

$$\begin{aligned}&E_\theta\{\bar{M}_a(\psi)\bar{M}_b(\psi)\} + E_\theta\{\bar{M}_{a/b}(\psi)\} \\ &\quad = \Delta_{ab} - \rho_a\rho_b + \rho_{a/r}\sigma_{bc}\lambda^{rc} + \sigma_{ac}\rho_{b/r}\lambda^{cr} - \beta_{ba} + O(n^{-1}) \qquad (a, b = 1, \ldots, p),\end{aligned} \tag{7}$$

where $\bar{M}_{a/b}(\psi) = \partial\bar{M}_a(\psi)/\partial\psi^b$. This simple formula facilitates calculation and comparison of the information bias for various adjustments that have been proposed. The formula also implies that differences to error of order $O(n^{-1})$ in the information bias for adjustment functions that satisfy equation (3) arise solely from differences in the quantities β_{ab}.

In the important case of the adjustment $B_a(\psi) = \rho_a\{\tilde{\theta}(\psi)\}$, it can be shown that $\beta_{ab} = \rho_{a/r}\sigma_{bc}\lambda^{rc} + O(n^{-1})$, and hence equation (6) reduces to

$$\zeta_a^b = (\Delta_{ac} - \rho_a\rho_c + \sigma_{ad}\rho_{c/r}\lambda^{dr})\lambda^{bc} + O(n^{-2}). \tag{8}$$

By using the formulae in Appendix A, it follows from equation (8) that

$$\begin{aligned}\zeta_a^b =&\sigma_{ac}\lambda^{br}\lambda^{cs}\nu^{tv}\nu^{uw}(\tfrac{1}{2}\,\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} - \tfrac{1}{2}\,\lambda_{st/r}\lambda_{uvw} + \lambda_{st/r}\lambda_{uv/w} \\ &+ \tfrac{1}{2}\,\lambda_{rtu}\lambda_{svw} - \lambda_{rtu}\lambda_{sv/w} - \lambda_{rt/u}\lambda_{svw} - \tfrac{1}{2}\,\lambda_{tu/r}\lambda_{svw} + \lambda_{tu/r}\lambda_{sv/w} + \lambda_{ru/t}\lambda_{sv/w}) \\ &- \sigma_{ac}\lambda^{br}\lambda^{cs}\nu^{tu}(\tfrac{1}{2}\,\lambda_{rstu} - \lambda_{rst/u} - \tfrac{1}{2}\,\lambda_{stu/r} + \lambda_{st/ru}) + O(n^{-2}).\end{aligned} \tag{9}$$

If ψ is scalar, then the adjusted profile score function can be used to obtain an adjusted profile log-likelihood function $\bar{M}(\psi)$ by integration. Possible definitions for $\bar{M}(\psi)$ include

$$\int_{\hat{\psi}}^{\psi}\{1+C_1^1(\xi)\}\,\bar{M}_1(\xi)\,\mathrm{d}\xi = M(\psi)+\int_{\hat{\psi}}^{\psi}\{B_1(\xi)+C_1^1(\xi)\,\bar{M}_1(\xi)\}\,\mathrm{d}\xi,$$

$$M(\psi)+\int_{\hat{\psi}}^{\psi}\{B_1(\xi)+C_1^1(\xi)\,M_1(\xi)\}\,\mathrm{d}\xi.$$

The score functions derived from these adjusted profile log-likelihoods have both bias and information bias of order $O(n^{-1})$. Unfortunately, if ψ is vector, then the integration approach to constructing $\bar{M}(\psi)$ is not generally feasible because $\bar{M}^{\dagger}_{a/b}(\psi)$ and $\bar{M}^{\dagger}_{b/a}(\psi)$ do not necessarily agree to error of order $O_p(n^{-1})$ for $a \neq b$. However, the construction of ζ_a^b does ensure symmetry in the expectation to error of order $O(n^{-1})$, i.e. $E_\theta\{\bar{M}^{\dagger}_{a/b}(\psi)\} = E_\theta\{\bar{M}^{\dagger}_{b/a}(\psi)\} + O(n^{-1})$. As a result, it is possible to construct an adjusted profile likelihood $\bar{M}(\psi)$ such that $\partial\bar{M}(\psi)/\partial\psi^a = \bar{M}^{\dagger}_a(\psi) + O_p(n^{-1})$; see Stern (1994).

If the additive adjustment $B_a(\psi)$ is invariant under reparameterizations of the form $(\psi, \phi) \rightarrow \{\psi, \eta(\psi, \phi)\}$, then β_{ab} and the approximation on the right-hand side of equation (6) are also parameterization invariant. In particular, $B_a(\psi) = \rho_a\{\tilde{\theta}(\psi)\}$ is invariant under such reparameterizations.

3. INFORMATION BIAS OF SPECIFIC ADJUSTMENTS

In this section, the information bias of the adjusted profile score function is calculated for additive adjustments proposed by Cox and Reid (1987), Barndorff-Nielsen (1983, 1994) and DiCiccio and Stern (1993). In each case, the function ζ_a^b required for equation (4) is determined so that the information bias is reduced to order $O(n^{-1})$.

3.1. *Case 1: Conditional Profile Likelihood of Cox and Reid (1987)*

In the case where ψ is orthogonal to the nuisance parameters, Cox and Reid (1987) considered adjusting the profile log-likelihood function according to equation (1) with

$$B(\psi) = -\frac{1}{2}\log\left(\frac{\det[-L_{\phi\phi}\{\tilde{\theta}(\psi)\}]}{\det\{-L_{\phi\phi}(\hat{\theta})\}}\right), \qquad (10)$$

where $L_{\phi\phi}(\theta)$ is the $q \times q$ matrix of second-order partial derivatives of $L(\theta)$ taken with respect to the nuisance parameters. The associated additive adjustment to the profile score function, $B_a(\psi) = \partial B(\psi)/\partial\psi^a$, satisfies equation (3) and has

$$\beta_{ab} = -\tfrac{1}{2}\lambda^{ij}\lambda_{abij} + \tfrac{1}{2}\lambda^{ik}\lambda^{jl}(\lambda_{abi}\lambda_{jkl} + \lambda_{aij}\lambda_{bkl}) + O(n^{-1}).$$

Therefore, the information bias given by equation (7) is

$$\lambda^{ij}\lambda_{abi/j} - \lambda^{ik}\lambda^{jl}\lambda_{abi}\lambda_{jk/l} + O(n^{-1}). \qquad (11)$$

Hence, although the Cox and Reid (1987) adjustment reduces the expectation of the score function to order $O(n^{-1})$, it does not generally reduce the information bias to order $O(n^{-1})$. However, if $\bar{M}_a^\dagger(\psi)$ is constructed using

$$\zeta_a^b = \lambda^{bc}\lambda^{ij}\lambda_{aci/j} - \lambda^{bc}\lambda^{ik}\lambda^{jl}\lambda_{aci}\lambda_{jk/l} + O(n^{-2}), \tag{12}$$

then both the bias and the information bias are of order $O(n^{-1})$. Some examples where the information bias (11) is automatically of order $O(n^{-1})$ are studied in Section 4. Section 5 concerns choosing an orthogonal parameterization that reduces the information bias.

3.2. *Case 2: Modified Profile Likelihood of Barndorff-Nielsen (1983)*

Barndorff-Nielsen (1983) proposed adjusting the profile log-likelihood function by adding

$$B(\psi) = -\frac{1}{2}\log\left(\frac{\det[-L_{\phi\phi}\{\tilde{\theta}(\psi)\}]}{\det\{-L_{\phi\phi}(\hat{\theta})\}}\right) + \log\left|\frac{\partial\hat{\phi}}{\partial\tilde{\phi}(\psi)}\right| \tag{13}$$

to $M(\psi)$, where $\tilde{\phi}(\psi)$ is regarded as a function of ψ, $\hat{\psi}$, $\hat{\phi}$ and an exact or approximate ancillary statistic A. It can be shown that $B_a(\psi)$ satisfies equation (3) and that

$$\begin{aligned}\beta_{ab} = &-\sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{ij}(\tfrac{1}{2}\lambda_{rsij} - \lambda_{rsi/j}) + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{ik}\nu^{jl}(\tfrac{1}{2}\lambda_{rsi}\lambda_{jkl} + \tfrac{1}{2}\lambda_{rij}\lambda_{skl} \\ &- \lambda_{rsi}\lambda_{jk/l} - \lambda_{rij}\lambda_{sk/l} - \lambda_{skl}\lambda_{ri/j} + \lambda_{ri/j}\lambda_{sl/k}) + O(n^{-1}).\end{aligned} \tag{14}$$

Substitution of equation (14) into equation (7) shows that the information bias of the adjusted profile score function associated with the modified profile likelihood of Barndorff-Nielsen (1983) is already of order $O(n^{-1})$. In deriving equation (14), the ancillarity of A is used to show that $L_{rs}(\hat{\theta}; \hat{\theta}, A) = \lambda_{rs}(\hat{\theta}) + O_p(n^{1/2})$, and hence $\partial l_{rs}(\hat{\theta})/\partial\hat{\theta}^t$ is of order $O_p(n^{1/2})$.

3.3. *Case 3: Barndorff-Nielsen (1994) Approximation to Modified Profile Likelihood*

For the case where ψ is scalar, Barndorff-Nielsen (1994) considered an approximation to adjustment (13) of the form

$$B(\psi) = \int_{\hat{\psi}}^{\psi} h\{\tilde{\theta}(\xi)\}\,\mathrm{d}\xi + \frac{1}{2}\log\left(\frac{\det[-\lambda_{\phi\phi}\{\tilde{\theta}(\psi)\}]}{\det\{-\lambda_{\phi\phi}(\hat{\theta})\}}\right) - \frac{1}{2}\log\left(\frac{\det[-L_{\phi\phi}\{\tilde{\theta}(\psi)\}]}{\det\{-L_{\phi\phi}(\hat{\theta})\}}\right), \tag{15}$$

where

$$h(\theta) = \sigma_{11}\lambda^{1r}\nu^{st}(\lambda_{rs/t} - \tfrac{1}{2}\lambda_{st/r}) = \sigma_{11}\lambda^{1r}\nu^{ij}(\lambda_{ri/j} - \tfrac{1}{2}\lambda_{ij/r})$$

and $\lambda_{\phi\phi}$ is the $q \times q$ submatrix of (λ_{rs}) corresponding to the nuisance parameters. Like adjustment (13), adjustment (15) is invariant under parameter transformations that preserve ψ, and the associated adjustment $B_1(\psi) = \partial B(\psi)/\partial\psi$ to $M_1(\psi)$ is also parameterization invariant. It can be shown that $B_1(\psi)$ satisfies equation (3) and that

$$\beta_{11} = -\sigma_{11}^2\lambda^{1r}\lambda^{1s}\nu^{tu}(\tfrac{1}{2}\lambda_{rstu} - \lambda_{rt/su}) + \sigma_{11}^2\lambda^{1r}\lambda^{1s}\nu^{tv}\nu^{uw}(\tfrac{1}{2}\lambda_{rst}\lambda_{uvw} - \tfrac{1}{2}\lambda_{rst}\lambda_{uw/v}$$
$$- \lambda_{rt/s}\lambda_{uv/w} + \tfrac{1}{2}\lambda_{rt/s}\lambda_{uw/v} + \tfrac{1}{2}\lambda_{rtu}\lambda_{svw} - \lambda_{tu/r}\lambda_{sv/w}) + O(n^{-1}).$$

The information bias given by equation (7) is then

$$\sigma_{11}^2\lambda^{1r}\lambda^{1s}\nu^{tu}(\lambda_{rst/u} - \lambda_{rt/su}) - \sigma_{11}^2\lambda^{1r}\lambda^{1s}\nu^{tv}\nu^{uw}(\lambda_{rst}\lambda_{uv/w} - \tfrac{1}{2}\lambda_{rst}\lambda_{uw/v} - \lambda_{rt/s}\lambda_{uv/w}$$
$$+ \tfrac{1}{2}\lambda_{rt/s}\lambda_{uw/v} + \lambda_{rtu}\lambda_{sv/w} + \lambda_{rt/u}\lambda_{svw} - \lambda_{rt/u}\lambda_{sw/v} - \lambda_{tu/r}\lambda_{sv/w}) + O(n^{-1}). \qquad (16)$$

Therefore, the appropriate form of ζ_1^1 for constructing $\bar{M}_1^\dagger(\psi)$ is obtained from expression (16) on multiplying by λ^{11}, which replaces σ_{11}^2 with σ_{11} and changes the error to $O(n^{-2})$. As for the Cox and Reid (1987) adjustment, Barndorff-Nielsen's (1994) adjustment reduces the bias of the profile score function to order $O(n^{-1})$, but it does not generally reduce the information bias to the same order. This observation arises from the fact that adjustment (15) is only a first-order approximation to adjustment (13); however, expression (16) can be used to construct a second-order approximation to adjustment (13) that would then reduce information bias.

3.4. *Case 4: Adjustment of DiCiccio and Stern (1993)*

In cases where $Y_1, \ldots, Y_n$ are independent and ψ is possibly a vector parameter of interest, DiCiccio and Stern (1993) considered the additive adjustment to the profile log-likelihood function

$$B(\psi) = -\frac{1}{2}\log\left(\frac{\det[-U_{\phi\phi}\{\tilde{\theta}(\psi)\}]}{\det\{-U_{\phi\phi}(\hat{\theta})\}}\right) + \frac{1}{2}\log\left(\frac{\det[-L_{\phi\phi}\{\tilde{\theta}(\psi)\}]}{\det\{-L_{\phi\phi}(\hat{\theta})\}}\right), \qquad (17)$$

where $U_{\phi\phi}(\theta)$ is the $q \times q$ matrix with elements

$$U_{ij}(\theta) = \sum_{m=1}^{n} L_i(\theta;\ Y_m)\, L_j(\theta;\ Y_m),$$

$L_i(\theta;\ Y_m) = \partial L(\theta;\ Y_m)/\partial\theta^i$, and $L(\theta;\ Y_m)$ is the log-likelihood function for θ based on the single observation Y_m. Adjustment (17) has the advantage of not requiring calculation of expected information. In the context of an exponential family, this adjustment function coincides with the function proposed by Barndorff-Nielsen and Cox (1979) when θ is the canonical parameterization. For adjustment (17), $B_a(\psi)$ satisfies equation (3), and

$$\beta_{ab} = \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}(\tfrac{1}{2}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} + \tfrac{3}{2}\lambda_{rtu}\lambda_{svw} - 2\lambda_{rtu}\lambda_{sv/w} - 2\lambda_{rt/u}\lambda_{svw}$$
$$+ 2\lambda_{rt/u}\lambda_{sv/w}) + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}(\tfrac{1}{2}\lambda_{rstu} + \lambda_{rst/u} + \lambda_{rt,su}) + O(n^{-1}),$$

where $\lambda_{rs,tu} = E_\theta\{\Sigma\, L_{rs}(\theta;\ Y_m)\, L_{tu}(\theta;\ Y_m)\}$ and $L_{rs}(\theta;\ Y_m) = \partial^2 L(\theta;\ Y_m)/\partial\theta^r\partial\theta^s$. Thus, the information bias given by equation (7) is

$$-\sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}(\lambda_{rtu}\lambda_{svw} - \lambda_{rtu}\lambda_{sv/w} - \lambda_{rt/u}\lambda_{svw} + 2\lambda_{rt/u}\lambda_{sv/w} - \lambda_{rt/u}\lambda_{sw/v})$$
$$-\sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\lambda_{rt,su} + O(n^{-1}).$$

Although adjustment (17) reduces the bias of the profile score to $O(n^{-1})$, it does not generally reduce the information bias to the same order. However, it follows from equation (6) that constructing $\bar{M}_a^\dagger(\psi)$ with

$$\begin{aligned}\zeta_a^b = &- \sigma_{ac}\lambda^{br}\lambda^{cs}\nu^{tv}\nu^{uw}(\lambda_{rtu}\lambda_{svw} - \lambda_{rtu}\lambda_{sv/w} - \lambda_{rt/u}\lambda_{svw} + 2\lambda_{rt/u}\lambda_{sv/w} - \lambda_{rt/u}\lambda_{sw/v}) \\ &- \sigma_{ac}\lambda^{br}\lambda^{cs}\nu^{tu}\lambda_{rt,su} + O(n^{-1})\end{aligned}$$

reduces the information bias to order $O(n^{-1})$.

4. EXAMPLES

4.1. *Example 1: Exponential Family*

Suppose that the log-likelihood function for θ based on $Y = (Y_1, \ldots, Y_n)$ is of the form

$$L(\theta) = g^a(\psi)\, T_a(Y) + \phi^i\, T_i(Y) + D(\theta),$$

up to an additive constant possibly depending on Y. For this model, no ancillary statistic is necessary for the modified profile likelihood of Barndorff-Nielsen (1983), since the maximum likelihood estimator is minimal sufficient. The additive adjustment (13) to the profile log-likelihood function is

$$B(\psi) = \frac{1}{2}\log\left(\frac{\det[-D_{\phi\phi}\{\tilde{\theta}(\psi)\}]}{\det\{-D_{\phi\phi}(\hat{\theta})\}}\right),$$

where $D_{\phi\phi}(\theta)$ is the $q \times q$ matrix of second-order partial derivatives of $D(\theta)$ taken with respect to the nuisance parameters; see also Barndorff-Nielsen and Cox (1979). This version of $B(\psi)$ has the property that $\partial B(\psi)/\partial\psi^a = \rho_a\{\tilde{\theta}(\psi)\}$ $(a = 1, \ldots, p)$, so the additive adjustment to the profile score function derived from the modified profile likelihood coincides with the adjustment of McCullagh and Tibshirani (1990). In particular, the additive adjustment $B_a(\psi) = \rho_a\{\tilde{\theta}(\psi)\}$ to the profile score reduces not only the bias but also the information bias, to order $O(n^{-1})$. This property is also apparent from equation (9); since $L_{ir}(\theta)$ is non-random in this case, $\lambda_{ir/s} = \lambda_{irs}$, and hence it follows from equation (9) that ζ_a^b is of order $O(n^{-2})$. Firth (1993) discussed bias reduction of the maximum likelihood estimator in this context and he pointed out that here $B_a(\psi) = \rho_a\{\tilde{\theta}(\psi)\}$ is an exact differential. Furthermore, when ψ is scalar, approximation (15) to adjustment (13) is exact, and it simplifies to

$$B(\psi) = \int_{\hat{\psi}}^{\psi} \rho_1\{\tilde{\theta}(\xi)\}\, \mathrm{d}\xi.$$

For this example, it is possible to reparameterize the model so that the nuisance parameters are orthogonal to ψ by transforming ϕ to η, where $\eta^i = E_\theta\{T_i(Y)\}$. When this orthogonal parameterization is used, the Cox and Reid (1987) adjustment (10) agrees with adjustment (13) for the modified profile likelihood. It can be shown that $\lambda_{abi} = 0$ for any orthogonal parameterization in this case, and thus, by expression (11), using adjustment (10) with any orthogonal parameterization reduces both the bias and the information bias to order $O(n^{-1})$. Any orthogonal nuisance parameters in this example are second order stable, i.e. $\tilde{\phi}(\psi)$ differs from $\hat{\phi}$ by order $O_p(n^{-3/2})$. See Section 5. This stability implies that adjustment functions (10) and (13) agree to error of order $O_p(n^{-3/2})$, and thus it explains why the information bias achieved by using adjustment (10) is reduced to order $O(n^{-1})$ in this case.

If the observations are independent and the log-likelihood function for θ based on the single observation Y_m is

$$L(\theta;\; Y_m) = g^a(\psi)\, T^m_a(Y_m) + \phi^i\, T^m_i(Y_m) + D^m(\theta),$$

then the adjusted profile score function obtained by using adjustment (17) of DiCiccio and Stern (1993) has information bias

$$-\sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\sum D^m_{rt}D^m_{su} + O(n^{-1}),$$

where $D^m_{rs} = \partial^2 D^m(\theta)/\partial\theta^r\partial\theta^s$. Use of adjustment (17) does not reduce the information bias to order $O(n^{-1})$ in general. However, if $D^{m_1}(\theta) = D^{m_2}(\theta)$ $(m_1, m_2 = 1, \ldots, n)$, as occurs when the observations are identically distributed, then $D^m_{ri}(\theta) = n^{-1}\lambda_{ri}$, and use of adjustment (17) does reduce the information bias to order $O(n^{-1})$.

4.2. *Example 2: Multiparameter Location Family*

Suppose that Y has density of the form

$$f_Y(y;\, \theta) = \prod_{m=1}^{n} f^m(y_m - \theta).$$

Then λ_{rs}, λ_{rst}, etc. do not depend on θ, and hence $\lambda_{rs/t}$, $\lambda_{rst/u}$, etc. all vanish. In this case, ρ_a is constant, and for the additive adjustment $B_a(\psi) = \rho_a$ the information bias is of order $O(1)$. It follows from equation (9) that $\bar{M}^{\dagger}_a(\psi)$, constructed by using

$$\zeta^b_a = \tfrac{1}{2}\sigma_{ac}\lambda^{br}\lambda^{cs}\nu^{tv}\nu^{uw}(\lambda_{rst}\lambda_{uvw} + \lambda_{rtu}\lambda_{svw}) - \tfrac{1}{2}\sigma_{ac}\lambda^{br}\lambda^{cs}\nu^{tu}\lambda_{rstu} + O(n^{-2}),$$

has information bias of order $O(n^{-1})$.

By using the usual ancillary statistic for this model, adjustment (13) which yields the Barndorff-Nielsen (1983) modified profile likelihood is

$$B(\psi) = -\frac{1}{2}\log\left(\frac{\det[-L_{\phi\phi}\{\tilde{\theta}(\psi)\}]}{\det\{-L_{\phi\phi}(\hat{\theta})\}}\right), \qquad (18)$$

where $\theta = (\psi, \phi)$. Note that adjustment (18) coincides with the Cox and Reid (1987) adjustment (10) obtained by using the same parameterization $\theta = (\psi, \phi)$, although in this case the parameters ψ and ϕ are not generally orthogonal. The adjusted profile score function obtained by using adjustment (18) has both bias and information bias of order $O(n^{-1})$. When ψ is scalar, adjustment (15) agrees with adjustment (13), and hence coincides with adjustment (18).

When adjustment (17) of DiCiccio and Stern (1993) is used to adjust the profile log-likelihood function, the resulting score function has information bias

$$-\sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}\lambda_{rtu}\lambda_{svw}-\sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}\lambda_{rt,su}+O(n^{-1}).$$

In particular, use of adjustment (17) does not typically reduce the information bias to order $O(n^{-1})$. However, $\bar{M}_a^{\dagger}(\psi)$, constructed by using

$$\zeta_a^b=-\sigma_{ac}\lambda^{br}\lambda^{cs}\nu^{tv}\nu^{uw}\lambda_{rtu}\lambda_{svw}-\sigma_{ac}\lambda^{br}\lambda^{cs}\nu^{tu}\lambda_{rt,su}+O(n^{-2}),$$

has information bias of order $O(n^{-1})$.

4.3. *Example 3: Location–Scale Family*

Suppose that the density of Y is

$$f_Y(y;\,\theta)=\psi^{-n}\prod_{m=1}^{n}f^m\{(y_m-\phi)/\psi\},$$

where ϕ is q dimensional. As for the previous example, λ_{rs}, λ_{rst}, etc. do not depend on ϕ, so that $\lambda_{rs/i}$, $\lambda_{rst/i}$, etc. all vanish. In this case, equation (9) yields

$$\begin{aligned}\zeta_1^1=&\tfrac{1}{2}\sigma_{11}\lambda^{1r}\lambda^{1s}\nu^{tv}\nu^{uw}(\lambda_{rst}\lambda_{uvw}+\lambda_{rtu}\lambda_{svw})-\tfrac{1}{2}\lambda^{1r}\nu^{su}\nu^{tv}(\lambda_{rst}\lambda_{uv/1}+\lambda_{rs/1}\lambda_{tuv})\\&-\tfrac{1}{2}\sigma_{11}\lambda^{1r}\lambda^{1s}\nu^{tu}\lambda_{rstu}+\tfrac{1}{2}\lambda^{1r}\nu^{st}\lambda_{rst/1}+O(n^{-2}).\end{aligned}$$

When the Barndorff-Nielsen (1994) adjustment (15) is used, it follows from expression (16) that the information bias is reduced to order $O(n^{-1})$. If the underlying density functions f^m ($m=1,\ldots,n$) are symmetric about the origin, then ψ and ϕ are orthogonal, and use of the Cox and Reid (1987) adjustment (10) also reduces the information bias to order $O(n^{-1})$.

For adjustment (17) of DiCiccio and Stern (1993), the information bias is

$$-\sigma_{11}^2\lambda^{1r}\lambda^{1s}\nu^{tv}\nu^{uw}\lambda_{rtu}\lambda_{svw}-\sigma_{11}^2\lambda^{1r}\lambda^{1s}\nu^{tu}\lambda_{rt,su}+O(n^{-1}).$$

If the underlying densities are symmetric about the origin, so that ψ and ϕ are orthogonal, the above expression for information bias simplifies to

$$-\lambda^{ik}\lambda^{jl}\lambda_{1ij}\lambda_{1kl}-\lambda^{ij}\lambda_{1i,1j}+O(n^{-1}).$$

As in the case of a multiparameter location family, the information bias of an adjusted profile score function based on adjustment (17) is typically of order $O(1)$.

4.4. *Example 4: Augmented Exponential Family*

Suppose that the log-likelihood function for θ based on Y is of the form

$$L(\theta)=\phi^i T_i(\psi,\,Y)+D(\theta)+H(\psi,\,Y),$$

so that for fixed ψ the distribution of Y belongs to an exponential family with canonical parameter ϕ. In this case, $L_{ij}(\theta)$ is non-random, and hence $\lambda_{ij/r}=\lambda_{ijr}$, $\lambda_{ijr/s}=\lambda_{ijrs}$, etc. For the adjustment $B_a(\psi)=\rho_a\{\tilde{\theta}(\psi)\}$, equation (9) yields

$$\zeta_a^b = -\lambda^{br}\nu^{su}\nu^{tv}(\tfrac{1}{2}\lambda_{ars}\lambda_{tuv} - \tfrac{1}{2}\lambda_{as/r}\lambda_{tuv} + \lambda_{rs/t}\lambda_{auv} - \lambda_{rs/t}\lambda_{av/u}) + \lambda^{br}\nu^{st}(\lambda_{ars/t} - \lambda_{as/rt}) + O(n^{-2}). \tag{19}$$

In this case, an appropriate ancillary statistic is not readily available, and thus application of the modified profile likelihood of Barndorff-Nielsen (1983) is not straightforward. For a scalar parameter of interest, the Barndorff-Nielsen (1994) adjustment (15), which does not require that an ancillary statistic be specified, simplifies to

$$B(\psi) = \int_{\hat{\psi}}^{\psi} \rho_1\{\tilde{\theta}(\xi)\}\,\mathrm{d}\xi,$$

as for a full exponential family. When this adjustment is used, the information bias is of order $O(1)$, and the appropriate ζ_1^1 for constructing $\bar{M}_1^\dagger(\psi)$ is given by equation (19).

If $E_\theta\{\partial T_i(\psi, Y)/\partial\psi^a\} = -\partial^2 D(\psi, \phi)/\partial\psi^a\partial\phi^i$, then the parameters ψ and ϕ are orthogonal. In particular, this occurs when $T_i(\psi, y) = a(\psi)\,T_i(y)$ and $D(\psi, \phi) = a(\psi)\,D(\phi)$, as happens for a generalized linear model with dispersion parameter ψ; see McCullagh and Nelder (1989). For inference about ψ, it follows from formula (9) that ζ_a^b is of order $O(n^{-2})$. When ψ is a scalar, the information bias obtained using either the Cox and Reid (1987) adjustment (10) or the Barndorff-Nielsen (1994) adjustment (15) is of order $O(n^{-1})$. If the observations $Y_1, \ldots, Y_m$ are independent and the log-likelihood contribution from the observation Y_m is

$$L(\theta;\, Y_m) = a(\psi)\phi^i\, T_i^m(Y_m) + a(\psi)\,D^m(\phi) + H^m(\psi,\, Y_m),$$

then the information bias arising from use of adjustment (17) proposed by DiCiccio and Stern (1993) is

$$\lambda^{ij}\lambda_{ai,bj} - \lambda^{ik}\lambda^{jl}\lambda_{aij}\lambda_{bkl} + O(n^{-1}).$$

Thus, as for a full exponential family, adjustment function (17) does not typically reduce the information bias to order $O(n^{-1})$. However, unlike the case for the full exponential family, even if the observations are identically distributed, the information bias generally remains of order $O(1)$.

5. CHOICE OF ORTHOGONAL PARAMETERIZATION

The difference between the constrained and overall maximum likelihood estimators of the nuisance parameter ϕ is generally of order $O_p(n^{-1/2})$. The difference is reduced to order $O_p(n^{-1})$ if and only if $\lambda_{ai} = 0$, i.e. if and only if the parameters ψ and ϕ are orthogonal. However, there are numerous possible orthogonal parameterizations of any particular parametric family. In fact, if (ψ, ϕ) is an orthogonal parameterization, then ψ and η are orthogonal parameters if and only if $\eta = g^{-1}(\phi)$ for some smooth invertible function g.

The information bias of the conditional profile likelihood of Cox and Reid (1987) for any given orthogonal parameterization is given by expression (11). If we consider an orthogonal reparameterization $\{\psi, g^{-1}(\phi)\}$ and set the corresponding expression for the information bias to 0, the resulting differential equation can be used to characterize those orthogonal parameterizations for which the Cox and Reid (1987) conditional profile likelihood has information bias of order $O(n^{-1})$. Thus, information bias can serve as an objective criterion for choosing between possible orthogonal parameterizations. Use of this criterion is illustrated in the following example.

5.1. *Example 5: Ratio of Exponential Means*

Suppose that $Y_m = (Y_{1,m}, Y_{2,m})$ $(m = 1, \ldots, n)$, where $Y_{1,m}$ and $Y_{2,m}$ are independent and exponentially distributed random variables having means ϕ and $\psi\phi$ respectively. Cox and Reid (1987) showed that ψ and $\phi\sqrt{\psi}$ are orthogonal parameters. Therefore, ψ and η are orthogonal parameters if and only if $\eta = g^{-1}(\phi\sqrt{\psi})$. When the parameterization (ψ, η) is used, the information bias of the adjusted profile score function based on the Cox and Reid (1987) adjustment (10) is

$$\frac{g''(\eta)\,g(\eta) - g'(\eta)^2}{\{2\psi\,g'(\eta)\}^2} + O(n^{-1}).$$

Clearly, the information bias is generally of order $O(1)$, but it can be reduced to order $O(n^{-1})$ if $\eta = k_1 \log(\phi\sqrt{\psi}) + k_2$ for constants $k_1 \neq 0$ and k_2. Cox and Reid (1993) noted that, for this choice of orthogonal parameters, their conditional profile likelihood coincides with the logarithm of the marginal likelihood function obtained from the F-distribution of the pivot $\psi\Sigma_m\, Y_{1,m}/\Sigma_m\, Y_{2,m}$.

5.2. *Comparisons with Cox and Reid (1989)*

In the case of a scalar parameter of interest, Cox and Reid (1989) discussed the possibility of finding an orthogonal parameterization under which the difference between $\tilde{\phi}(\psi)$ and $\hat{\phi}$ is of order $O_p(n^{-3/2})$, instead of order $O_p(n^{-1})$. If such an orthogonal parameterization is employed, the second term on the right-hand side of equation (13) is of a negligible order; see Barndorff-Nielsen (1994). In such cases, the Cox and Reid (1987) conditional profile likelihood is equivalent to the Barndorff-Nielsen (1983) modified profile likelihood to error of order $O_p(n^{-3/2})$, and therefore it has information bias of order $O(n^{-1})$. Although Cox and Reid (1989) showed that such orthogonal parameterizations are not generally available, they provided a criterion for choosing between orthogonal parameterizations. They suggested choosing the nuisance parameters η^i to be q functionally independent solutions to the differential equation

$$\lambda^{jk}\lambda_{11j}\,\frac{\partial\eta^i}{\partial\phi^k} = \text{constant}.$$

Cox and Reid (1989) pointed out that such an orthogonal parameterization would make 'the quadratic variation of $\tilde{\phi}(\psi)$ with ψ as free of nuisance parameter effects as possible'. By using overbars to indicate expected values of the derivatives of the log-likelihood function with respect to the parameterization (ψ, η), the Cox and Reid (1989) criterion can be written as

$$\lambda^{kl}\lambda_{11k}\,\frac{\partial\eta^i}{\partial\phi^l} = \bar{\lambda}^{ij}\bar{\lambda}_{11j} = \text{constant}.$$

By differentiating this equation with respect to η^j, this criterion is seen to produce the system of equations $\bar{H}^i_j = 0$, where

$$\bar{H}^i_j = \bar{\lambda}^{ik}\bar{\lambda}_{11k/j} - \bar{\lambda}^{il}\bar{\lambda}^{km}\bar{\lambda}_{11k}\bar{\lambda}_{lm/j}.$$

A comparison with expression (11) shows that the information bias for the Cox and Reid (1987) conditional profile likelihood in the orthogonal parameterization (ψ, η) is the trace of the matrix $(\bar{H}^i_j)$. Therefore, the Cox and Reid (1989) criterion for choosing between orthogonal parameterizations is closely related to a criterion that chooses an orthogonal parameterization to reduce the information bias of the resulting conditional profile likelihood to order $O(n^{-1})$. In the case of a single nuisance parameter, the two criteria are equivalent. For multiple nuisance parameters, the Cox and Reid (1989) criterion requires all the diagonal elements of $(\bar{H}^i_j)$ to be 0, whereas the criterion based on the reduction of information bias requires only that these elements sum to 0. Thus, any orthogonal parameterization which satisfies the Cox and Reid (1989) criterion will provide a conditional profile likelihood with lower order information bias, although the converse is generally not true. The choice of a parameterization on the basis of reducing information bias provides a compelling justification for the Cox and Reid (1989) procedure. Moreover, since the criterion based on reducing information bias is less strict than the criterion of Cox and Reid (1989), it may yield a solution in a broader class of problems.

ACKNOWLEDGEMENTS

The helpful comments of the Editor and the referees are gratefully acknowledged.

APPENDIX A

Standard calculations show that

$$\begin{aligned}\Delta_{ab} = & -\sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tu}(\tfrac{1}{2}\lambda_{rstu} - \lambda_{rst/u} - \tfrac{1}{2}\lambda_{stu/r} - \tfrac{1}{2}\lambda_{rtu/s} + \lambda_{st/ru} + \lambda_{rt/su}) \\ & + \sigma_{ac}\sigma_{bd}\lambda^{cr}\lambda^{ds}\nu^{tv}\nu^{uw}(\tfrac{1}{2}\lambda_{rst}\lambda_{uvw} - \lambda_{rst}\lambda_{uv/w} - \tfrac{1}{2}\lambda_{rt/s}\lambda_{uvw} - \tfrac{1}{2}\lambda_{st/r}\lambda_{uvw} \\ & + \lambda_{rt/s}\lambda_{uv/w} + \lambda_{st/r}\lambda_{uv/w} + \tfrac{1}{2}\lambda_{rtu}\lambda_{svw} - \lambda_{rtu}\lambda_{sv/w} - \lambda_{rt/u}\lambda_{svw} \\ & - \tfrac{1}{2}\lambda_{rtu}\lambda_{vw/s} - \tfrac{1}{2}\lambda_{tu/r}\lambda_{svw} + \lambda_{rt/u}\lambda_{vw/s} + \lambda_{rt/u}\lambda_{sw/v} + \lambda_{tu/r}\lambda_{sv/w} \\ & + \tfrac{1}{4}\lambda_{rtv}\lambda_{suw} - \tfrac{1}{2}\lambda_{rtv}\lambda_{su/w} - \tfrac{1}{2}\lambda_{rt/v}\lambda_{suw} + \lambda_{rt/v}\lambda_{su/w}) + O(n^{-1}),\end{aligned}$$

$$\begin{aligned}\rho_{a/r} = & \sigma_{ab}\lambda^{bs}\nu^{tv}\nu^{uw}(\tfrac{1}{2}\lambda_{st/r}\lambda_{uvw} - \lambda_{st/r}\lambda_{uv/w} + \tfrac{1}{2}\lambda_{tu/r}\lambda_{svw} - \lambda_{tu/r}\lambda_{sv/w}) \\ & - \sigma_{ab}\lambda^{bs}\nu^{tu}(\tfrac{1}{2}\lambda_{stu/r} - \lambda_{st/ru}).\end{aligned}$$

REFERENCES

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

——— (1994) Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *J. R. Statist. Soc.* B, **56**, 125–140.

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

Barndorff-Nielsen, O. E. and Cox, D. R. (1979) Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. R. Statist. Soc.* B, **41**, 279–312.

Bartlett, M. S. (1955) Approximate confidence intervals, III. A bias correction. *Biometrika*, **42**, 201–204.

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

——(1989) On the stability of maximum-likelihood estimators of orthogonal parameters. *Can. J. Statist.*, **17**, 229–233.

——(1992) A note on the difference between profile and modified profile likelihood. *Biometrika*, **79**, 408–411.

——(1993) A note on the calculation of adjusted profile likelihood. *J. R. Statist. Soc.* B, **55**, 467–471.

DiCiccio, T. J. and Stern, S. E. (1993) An adjustment to profile likelihood based on observed information. *Technical Report*. Department of Statistics, Stanford University, Stanford.

——(1994) Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods. *J. R. Statist. Soc.* B, **56**, 397–408.

Ferguson, H., Reid, N. and Cox, D. R. (1991) Estimating equations based on modified profile likelihood. In *Estimating Functions* (ed. V. P. Godambe), pp. 279–293. Oxford: Oxford University Press.

Firth, D. (1993) Bias reduction of maximum likelihood estimates. *Biometrika*, **80**, 27–38.

Godambe, V. P. (1960) An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.*, **31**, 1208–1211.

Levin, B. and Kong, F. (1990) Bartlett's bias correction to the profile score function is a saddlepoint correction. *Biometrika*, **77**, 219–221.

Liang, K.-Y. (1987) Estimating functions and approximate conditional likelihood. *Biometrika*, **74**, 695–702.

Lindsay, B. (1982) Conditional score functions: some optimality results. *Biometrika*, **69**, 503–512.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

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

Stern, S. E. (1994) Asymptotic corrections to adjusted profile likelihoods. *PhD Thesis*. Department of Statistics, Stanford University, Stanford.