A Note on the Relation Between Modified Profile Likelihood and the Cox-Reid Adjusted Profile Likelihood

Ole E. Barndorff-Nielsen; Peter McCullagh

Biometrika, Volume 80, Issue 2 (Jun., 1993), 321-328.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28199306%2980%3A2%3C321%3AANOTRB%3E2.0.CO%3B2-2

Biometrika (1993), **80**, 2, *pp.* 321–8
Printed in Great Britain

A note on the relation between modified profile likelihood and the Cox–Reid adjusted profile likelihood

BY OLE E. BARNDORFF-NIELSEN
Department of Theoretical Statistics, Aarhus University, DK 8000 Aarhus C, Denmark
AND PETER McCULLAGH
Department of Statistics, The University of Chicago, 5734 University Avenue, Chicago, Illinois 60637, U.S.A.

SUMMARY

An adjustment to the profile likelihood proposed by Cox & Reid (1987) when the parameters are orthogonal is shown to agree with modified profile likelihood in a number of instances in which the parameters are not orthogonal. Several examples of the phenomenon are given. In the case of composite transformation models with invariant parameter as the parameter of interest, it is shown that the Cox–Reid adjustment agrees with modified profile likelihood in a large-deviation sense if Lebesgue measure is right-invariant. Necessary and sufficient conditions for equivalence in the moderate-deviation sense of modified profile likelihood and the Cox–Reid adjusted profile likelihood are derived.

Some key words: Adjusted profile likelihood; Group transformation model; Invariant measure; Modified profile likelihood; Nuisance parameter.

1. INTRODUCTION

The purpose of this paper is to show that an adjustment to the profile likelihood proposed by Cox & Reid (1987) when the parameters are orthogonal agrees with modified profile likelihood (Barndorff-Nielsen, 1983, 1988) in a number of instances in which the parameters are not orthogonal. By agreement in this context we mean that the adjusted profile likelihood, the modified profile likelihood and any obvious marginal or conditional likelihood differ by terms of order $O(n^{-1})$. In this context, a distinction is drawn between agreement in a large-deviation sense and agreement in a moderate-deviation sense. In the former case the adjusted profile likelihood agrees with the marginal or conditional likelihood in an $O(1)$ region of the parameter space: in the latter case the two likelihoods agree on an $O(n^{-\frac{1}{2}})$ region of the parameter space.

Several examples of equivalence are given. In the case of composite transformation models in which the invariant parameter is the parameter of interest, the Cox–Reid adjustment is shown to agree with modified profile likelihood in the large-deviation sense if Lebesgue measure is right-invariant, and to agree in the moderate-deviation sense if the parameters are orthogonal. For location-scale models in which the location parameter is of interest, the Cox–Reid adjustment agrees with modified profile likelihood in the moderate-deviation sense, but not in the large-deviation sense.

Necessary and sufficient conditions for equivalence in the moderate-deviation sense of modified profile likelihood and the Cox–Reid adjusted profile likelihood are derived in § 6.

2. MODIFIED PROFILE LIKELIHOOD AND COX–REID ADJUSTED PROFILE LIKELIHOOD

Let ψ, of dimension d, be the parameter of interest, and χ, of dimension k, the nuisance parameter. We assume that $(\hat{\psi}, \hat{\chi}, a)$ is sufficient, where $\hat{\psi}, \hat{\chi}$ are maximum likelihood estimators, and a is asymptotically ancillary. The profile likelihood $L_p(\psi)$, the modified profile likelihood $L_{mp}(\psi)$ (Barndorff-Nielsen, 1983, 1988), and the adjusted profile likelihood $L_{ap}(\psi)$ considered by Cox & Reid (1987) are related by

$$L_{mp}(\psi) = D(\psi) L_{ap}(\psi) = D(\psi) |\hat{j}_{\chi\chi}(\psi, \hat{\chi}_\psi)|^{-\frac{1}{2}} L_p(\psi).$$

The factor $D(\psi)$ may be expressed as

$$D(\psi) = |\partial\hat{\chi}/\partial\hat{\chi}_\psi| \tag{1}$$

for $\hat{\chi}_\psi$ considered as a function of $(\hat{\psi}, \hat{\chi}, a)$, or, equivalently, as

$$D(\psi) = \frac{|j_{\chi\chi}(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a)|}{|l_{\chi;\hat{\chi}}(\psi, \hat{\chi}_\psi; \hat{\psi}, \hat{\chi}, a)|}. \tag{2}$$

In (2) the dependence on $\hat{\psi}$, $\hat{\chi}$ and a of the observed information $j_{\chi\chi}$ has been made explicit; furthermore, $l_{\chi;\hat{\chi}}$ is the matrix of second-order mixed derivatives relative to χ and $\hat{\chi}$. Note that we always have

$$j_{\chi\chi}(\hat{\psi}, \hat{\chi}; \hat{\psi}, \hat{\chi}, a) = l_{\chi;\hat{\chi}}(\hat{\psi}, \hat{\chi}; \hat{\psi}, \hat{\chi}, a),$$

as follows on differentiating the likelihood equation for $\hat{\chi}$, and hence $D(\hat{\psi}) = 1$. The equivalence of the above-mentioned two versions of $D(\psi)$ is seen via differentiation of the likelihood equation for $\hat{\chi}_\psi$.

In order to investigate the conditions under which the Cox–Reid adjusted profile likelihood in a given parameterization is equivalent to the modified profile likelihood we examine conditions under which

$$D(\psi) = 1 \tag{3}$$

or more generally

$$D(\psi) = 1 + O(n^{-1}). \tag{4}$$

The Cox–Reid approximation is not invariant to the choice of nuisance parameter. Consequently the conditions under which (4) holds must also depend on the choice of nuisance parameter, although this dependence is not made explicit in the notation.

We say that the two approximations are equivalent to $O(n^{-1})$ in the large-deviation sense if (4) holds for $\psi - \hat{\psi} = O(1)$: if (4) holds for $\psi - \hat{\psi} = O(n^{-\frac{1}{2}})$ the two approximations are said to be equivalent in the weaker moderate-deviation sense.

If $\hat{\chi} = \hat{\chi}_\psi$ then (3) holds, as is obvious from (1). It is, furthermore, known that the more general condition of parameter orthogonality of ψ and χ in the expected or observed information sense is sufficient to ensure (4) in the moderate-deviation sense. In such cases the factor $D(\psi)$, which is sometimes computationally awkward, can be discarded with no or little sacrifice of information in large samples. The examples presented in the following section show that in certain circumstances the stronger condition (3) is satisfied even in the absence of parameter orthogonality, by which we mean that $\hat{j}_{\psi\chi} = O_p(n^{\frac{1}{2}})$ or $i_{\psi\chi} = O(n^{\frac{1}{2}})$, not necessarily zero. In other words parameter orthogonality is sufficient, but not necessary, to justify discarding the factor $D(\psi)$ in the Cox–Reid adjusted profile likelihood approximation. It should be kept in mind, however, that without the factor $D(\psi)$ there is no longer parameterization invariance of the adjusted profile likelihood, and also that parameter orthogonality ensures only $O(n^{-1})$ accuracy.

3. Cases of equality of L_{mp} and L_{ap}

Example 1: *Location-scale model.* Let $f(.)$ be an arbitrary density function, possibly asymmetric, and let Y_i have a density of the form $\sigma^{-1}f((y_i-\mu)/\sigma)$. A standard example is the log-Weibull or extreme-value distribution in which $f(x)=\exp(x-e^x)$ for $-\infty< x<\infty$. The log likelihood for (μ, σ) in a simple random sample of size n is

$$\begin{aligned} l(\mu, \sigma) &= -n \log \sigma - \sum g((y_i-\mu)/\sigma) \\ &= n \log \sigma - \sum g((\hat{\mu}-\mu+a_i\hat{\sigma})/\sigma) \\ &= l(\mu, \sigma; \hat{\mu}, \hat{\sigma}, a), \end{aligned}$$

where $g=-\log f$, $a_i=(y_i-\hat{\mu})/\hat{\sigma}$ and $a=(a_1, \ldots, a_n)$ is ancillary.

With σ as the parameter of interest we find

$$l_{\mu\mu}(\mu, \sigma) = -\sigma^{-2} \sum g''(\sigma^{-1}(y_i-\mu)), \quad l_{\mu;\hat{\mu}}(\mu, \sigma) = \sigma^{-2} \sum g''(\sigma^{-1}(y_i-\mu)).$$

Hence in this case we have

$$l_{\mu;\hat{\mu}}(\mu, \sigma) = j_{\mu\mu}(\mu, \sigma),$$

which is even stronger than (3). Furthermore the parameters are not orthogonal even in the expected Fisher information sense unless $f(.)$ is symmetric, or, slightly more generally, unless $\int yg''(y)f(y)\,dy=0$. In the Weibull example we have $i_{\mu\sigma}=n(\pi^2/6-1)/\sigma^3$.

Example 2: *Translation models and residual likelihood.* Example 1 is a special case of the following linear regression model. Suppose that the random vector Y has independent components Y_i with location parameter μ_i and density of the form

$$f(y_i-\mu_i;\tau)$$

depending on a parameter τ, which need not be a scale parameter. Suppose, in addition, that the vector μ satisfies $\mu=X\beta$ in the usual notation for linear models in which X is a given $n\times p$ model matrix and β is an unknown parameter vector. The parameters are $(\beta;\tau)$ in which τ is taken to be the component of interest. Both components may be vector-valued. In the Normal-theory case the transformation from Y to $(\hat{\beta}, R)$, where $R=(I-X(X^TX)^{-1}X^T)Y$, is linear, and hence the Jacobian is a constant that depends on X. More generally, for non-Normal models, $\hat{\beta}$ is equivariant and R is invariant. It can be shown that, although the transformation is no longer linear, the Jacobian depends only on X. The joint density of $\hat{\beta}$ and R has the form

$$f(\hat{\beta}-\beta, R;\tau)J(X),$$

and the marginal likelihood based on R alone is

$$L_M(\tau; R) = \int_{R^p} L(\tau, \beta; Y)\,d\beta,$$

where $L(\tau, \beta; Y)$ is the joint likelihood based on Y; compare, for instance, Proposition 2.2 of Barndorff-Nielsen (1988). Laplace's approximation to the above integral gives

$$L_M(\tau; R) \simeq L(\tau, \hat{\beta}_\tau; Y)|j_{\beta\beta}(\tau, \hat{\beta}_\tau)|^{-\frac{1}{2}},$$

which is the Cox-Reid approximation in the (τ, β) parameterization, again derived without the benefit of orthogonality. In this example also, $D(\psi)=1$, so that the Cox-Reid and modified profile likelihoods are approximations to the marginal or residual likelihood based on R. For some related discussions, see Barndorff-Nielsen (1988) and Barndorff-Nielsen & Jupp (1988).

It is not necessary in this example that μ_i be the mean or median of Y_i. Any location parameter such as a specified quantile is equally satisfactory. That is, the argument is unaffected if we replace μ_i by the location parameter $\gamma_i = \mu_i + w(\tau)$, where $w(.)$ is a known function. If this new location parameter satisfies the regression model $\gamma = X\beta$, the Cox–Reid approximation in the (τ, β) parameterization is unaffected. The marginal likelihood based on R depends only on τ as before, and agrees with both the Cox–Reid approximation and the modified profile likelihood in the large-deviation sense. Note that in the Normal-theory case with $\mu_i = E(Y_i)$, the parameters are not orthogonal unless $w(\tau)$ is a constant.

4. General theory for composite transformation models

Examples 1 and 2 are special cases of composite transformation models in which ψ is the invariant parameter. The group G, whose elements we denote by g, is assumed to act freely on the sample space, that is $gy \neq y$ unless g is the identity in G. Its action on the parameter space has no effect on ψ. The model function is the form

$$p(y; g, \psi) = p(u; \hat{g}^{-1}g, \psi),$$

where y is the full data vector,

$$u = \hat{g}^{-1}y$$

is a maximal invariant, and p is a density with respect to invariant measure on the sample space. The group element g plays the role of the nuisance parameter χ, and $\hat{g}$ is any equivariant estimator satisfying $\hat{g}(gy) = g\hat{g}(y)$. For instance, $\hat{g}$ may be taken to be the maximum-likelihood estimator.

The marginal distribution of the maximal invariant depends only on the invariant parameter ψ. From Proposition 2.2 of Barndorff-Nielsen (1988), the marginal likelihood for ψ based on u is given by

$$L(\psi; u) = \int_G p(y; g, \psi)\, d\bar{\rho}(g), \tag{5}$$

where $\bar{\rho}$ denotes right-invariant measure on G. Assuming that $\bar{\rho}$ is absolutely continuous with density ρ relative to Lebesgue measure, Laplace approximation gives

$$L(\psi; u) = L_{ap}(\psi)\rho(\hat{g}_\psi)\{1 + O(n^{-1})\},$$

where $\hat{g}_\psi = \hat{\chi}_\psi$ is the maximum-likelihood estimator of χ for fixed ψ. This expression reduces to the Cox–Reid adjusted profile likelihood in the (ψ, g) parameterization if and only if $\rho(\hat{g}_\psi)$ is independent of ψ when terms of order $O(n^{-1})$ are ignored. From the Taylor expansion about $\hat{g}$

$$\rho(\hat{g}_\psi) = \rho(\hat{g}) + \rho'(\hat{g})(\hat{g}_\psi - \hat{g}) + \ldots$$

it follows that for that to happen either $\hat{g}_\psi - \hat{g} = O(n^{-1})$ or else $\rho'(\hat{g}) = 0$. The former condition is satisfied if the parameters are orthogonal as required by Cox & Reid. The second condition, that the right invariant measure be constant, is satisfied by the examples in § 3.

Note that, if $\rho(.)$ is constant, the Cox–Reid adjusted profile likelihood in the (τ, g) parameterization agrees with the marginal likelihood and the modified profile likelihood in the large-deviation sense. Parameter orthogonality guarantees approximate agreement only in the moderate-deviation sense.

5. A further example of equivalence

Example 3: *Equivalence without group structure.* In order to demonstrate that the phenomenon described in § 4 is not peculiar to composite transformation models, we present the following example in which group structure as usually understood is absent. Suppose that the observed random variables U, V are generated as follows

$$U \sim N_k(A_\psi \chi, I), \quad V \mid U \sim p(v; \psi, u),$$

in which A_ψ, depending on ψ, is a matrix with unit determinant, and $p(v; \psi, u)$ is an arbitrary density. Then we have

$$\hat{\chi}_\psi = A_\psi^{-1} U = A_\psi^{-1} A_{\hat{\psi}} A_{\hat{\psi}}^{-1} U = A_\psi^{-1} A_{\hat{\psi}} \hat{\chi}.$$

If $\det A_\psi = 1$ for all ψ then

$$D(\psi) = |\partial \hat{\chi}_\psi / \partial \hat{\chi}|^{-1} = 1,$$

so condition (3) is satisfied. Thus, again, the Cox–Reid adjusted profile likelihood in the (ψ, χ) parameterization coincides exactly with the modified profile likelihood even though the parameters are not in general orthogonal. Further, in this example the modified and adjusted profile likelihoods coincide with the conditional likelihood based on V given $U = u$.

One can in this case orthogonalize the parameters by re-defining the nuisance parameter as $\xi = A_\psi \chi$. The likelihood now factors into

$$L_{uv}(\xi, \psi) = L_u(\xi) L_{v|u}(\psi),$$

from which it is clear that only the second factor contributes to information on ψ (provided that ξ and ψ are variation independent).

6. General conditions for moderate-deviation equivalence

We denote the generic components of ψ and χ by $\psi^a, \psi^b, \ldots$ and $\chi^r, \chi^s, \ldots$ respectively. The leading term in the Taylor expansion of $d(\psi) = \log D(\omega)$ in ψ about $\hat{\psi}$ may be written in the form

$$d(\psi) \simeq (\psi - \hat{\psi})^a \hat{j}_{\chi\chi}^{rs} \{\hat{j}_{rs/a} - \hat{l}_{r;s/a} + (\hat{j}_{rs/t} - \hat{l}_{r;s/t}) \hat{\chi}^t_{\psi/a}\}$$

in which the subscript $/a$ denotes differentiation. As a convention, we assume that evaluation at $(\hat{\psi}, \hat{\chi})$, which is indicated by $\hat{}$, is the final operation in each term. Differentiation of the likelihood equation for $\hat{\chi}_\psi$, that is $l_r(\psi, \hat{\chi}_\psi) = 0$, with respect to ψ^a gives

$$l_{ra}(\psi, \hat{\chi}_\psi) + l_{rs}(\psi, \hat{\chi}_\psi) \hat{\chi}^s_{\psi/a} = 0.$$

Hence the leading term in $d(\psi)$ is

$$\begin{aligned} d(\psi) &\simeq (\psi - \hat{\psi})^a \hat{j}_{\chi\chi}^{rs} \{\hat{j}_{rs/a} - \hat{l}_{r;s/a} - \hat{j}_{\chi\chi}^{tu} \hat{l}_{ua} (\hat{j}_{rs/t} - \hat{l}_{r;s/t})\} \\ &= (\hat{\psi} - \psi)^a \hat{j}_{\chi\chi}^{rs} \{(\hat{l}_{ra})_{/s} - \hat{j}_{\chi\chi}^{tu} \hat{l}_{ua} (\hat{l}_{rs})_{/t}\} \\ &= (\hat{\psi} - \psi)^a \{(\hat{j}_{\chi\chi})^{rs} \hat{l}_{ra}\}_{/s}, \end{aligned}$$

where, by $(\hat{l}_{ra})_{/s}$, we mean $\partial l_{ra}(\hat{\psi}, \hat{\chi}; \hat{\psi}, \hat{\chi}, a)/\partial \hat{\chi}^s$, that is evaluation at $(\hat{\psi}, \hat{\chi})$ followed by differentiation with respect to $\hat{\chi}^s$ for a given value of the ancillary $A = a$.

For moderate deviations in which $\psi-\hat{\psi}=O(n^{-\frac{1}{2}})$, we conclude that $d(\psi)$ is of order $O(n^{-1})$ or smaller if and only if

$$\{(\hat{j}_{\chi\chi})^{rs}\hat{l}_{ra}\}_{/s}=O(n^{-\frac{1}{2}}), \tag{6}$$

or equivalently, if and only if

$$\{i^{rs}_{\chi\chi}i_{ra}\}_{/s}=O(n^{-\frac{1}{2}}), \tag{7}$$

where i denotes expected information.

In the location-scale example with μ as parameter of interest, this condition reduces to

$$\hat{j}^{-1}_{\sigma\sigma}\hat{j}_{\sigma\mu}=\text{const}+O(n^{-\frac{1}{2}}),$$

where the constant is a quantity that does not depend on $\hat{\sigma}$. The condition is fulfilled because both $\hat{j}_{\sigma\sigma}$ and $\hat{j}_{\sigma\mu}$ are of the form $\hat{\sigma}^{-2}$ times a function of the configuration ancillary.

7. AN EXAMPLE IN WHICH LEBESGUE MEASURE IS NOT RIGHT-INVARIANT

Example 4: *Location-scale problem with μ as parameter of interest.* Suppose that $Y_1,\ldots,Y_n$ are independent with density $\sigma^{-1}f((y-\mu)/\sigma)$. For fixed μ, let Σ_μ be the scale group that leaves the point μ invariant. Elements σ of Σ_μ have a representation in terms of the positive real numbers acting on the sample space as follows:

$$\sigma\circ y=\mu+\sigma(y-\mu).$$

The composition of two group elements corresponds to ordinary multiplication. There is a unique identity element $\sigma=1$, and an inverse element $1/\sigma$.

The left and right invariant measures on R^+ under Σ_μ are both $d\sigma/\sigma$, the same for each μ.

This is a transformation model, but not a composite transformation model with μ as invariant parameter, because the action of the group Σ_μ on the sample space depends on the index parameter. Consequently the maximal invariant under the group depends on μ, so it does not make sense to talk of the marginal likelihood for μ based on the invariant. Nevertheless, it is possible formally to use the integration formula (5), though the usual justification in terms of marginal likelihood based on a single statistic is no longer available. This line of argument gives a 'marginal' or integrated likelihood for μ in the form

$$L(\mu;y)=\int_0^\infty\frac{1}{\sigma^n}\prod_{i=1}^n f\left(\frac{y_i-\mu}{\sigma}\right)\frac{d\sigma}{\sigma}. \tag{8}$$

This expression has a mixture interpretation as the marginal density of Y, and a Bayesian interpretation as a marginal posterior for μ. In either case, the right-invariant measure represents an improper 'prior' or 'mixing distribution' for σ, which is independent of μ.

Laplace approximation to the logarithm of (8) gives

$$l(\mu;y)=l(\mu,\hat{\sigma}_\mu;y)-\tfrac{1}{2}\log|j_{\sigma\sigma}(\mu,\hat{\sigma}_\mu)|-\log\hat{\sigma}_\mu+O(n^{-1}). \tag{9}$$

In the Cox–Reid approximation with the (μ,σ) parameterization, the third term above is absent. In the modified profile likelihood, the third term is replaced by $\log D(\mu)$. Detailed specific calculations concerning the relation between $D(\mu)$ and $\hat{\sigma}_\mu$ give

$$\log D(\mu)=-\log\hat{\sigma}_\mu+\log\left\{\hat{\sigma}+\frac{l_{\mu\sigma;}(\mu,\hat{\sigma}_\mu)}{l_{\sigma;\hat{\sigma}}(\mu,\hat{\sigma}_\mu)}(\mu-\hat{\mu})\right\}. \tag{10}$$

Since

$$l_{\mu\sigma;}(\hat{\mu}, \hat{\sigma}) = -\hat{j}_{\mu\sigma}, \quad l_{\sigma;\sigma}(\hat{\mu}, \hat{\sigma}) = \hat{j}_{\sigma\sigma}, \quad \hat{\sigma}_\mu = \hat{\sigma} - \hat{j}_{\mu\sigma}(\mu - \hat{\mu})/\hat{j}_{\sigma\sigma} + O(n^{-1}),$$

it follows that $\log D(\mu) = O_p(n^{-1})$ for $\mu - \hat{\mu} = O(n^{-\frac{1}{2}})$. Thus, with error $O(n^{-1})$, the Cox–Reid and modified profile likelihood approximations are equivalent in the moderate-deviation sense, but they differ from the integrated likelihood (8) by $O(n^{-\frac{1}{2}})$.

A referee has pointed out that if σ is replaced by $\xi = \log \sigma$, the Cox–Reid approximation in the (μ, ξ) parameterization becomes $l(\mu, \hat{\xi}_\mu; y) - \frac{1}{2}\log |j_{\xi\xi}(\mu, \hat{\xi}_\mu)|$, which is then identical to the first three terms in (9). Thus the Cox–Reid approximation can be made to agree with either (8) or (9) by choice of parameterization.

If the observations are normally distributed with mean μ and variance σ^2, the integrated density (8) has the property that $(\bar{Y}, s^2)$ is minimal sufficient and s^2 is ancillary. However, $\bar{Y}$ and s^2 are not marginally independent. The conditional distribution of $\bar{Y}$ given s^2 is such that

$$\bar{Y} = \mu + t_{n-1}s/\sqrt{n},$$

where t_{n-1} is Student's distribution (Cox, 1975). More generally, if μ is replaced by the location parameter $\gamma = \mu + k\sigma$, a quantile specified by k, the sufficient statistics in (8) are unaffected and s^2 remains ancillary. The conditional density of $\bar{Y} - \gamma$ given s^2 is proportional to (8) expressed as a function of $\bar{y} - \gamma$ and s^2. Equivalently, the pivot $\sqrt{n}(\bar{Y} - \gamma)/s$ has the non-central Student's t distribution on $n-1$ degrees of freedom with non-centrality parameter $-k\sqrt{n}$. This pivot can be used to generate exact confidence intervals for γ (Fisher, 1931).

More generally, whether $f(.)$ is normal or not, the integrated likelihood (8) is proportional to the marginal density of the pivot $(\hat{\mu} - \mu)/\hat{\sigma}$ given the location-scale configuration vector. In other words, if we transform from the original Y to the joint pivot

$$P_1 = \frac{\hat{\mu} - \mu}{\hat{\sigma}}, \quad P_2 = \frac{\hat{\sigma}}{\sigma},$$

with the configurations vector $C = \{(Y_1 - \hat{\mu})/\hat{\sigma}, \ldots, (Y_n - \hat{\mu})/\hat{\sigma}\}$ as complementary statistic, the marginal density of P_1 given $C = c$ is proportional to (8). Thus, the exact conditional confidence density for μ in the sense of Fisher (1934) is obtained from the integrated likelihood (8) by normalization so that it integrates to one. One might, therefore, have expected the modified profile likelihood to agree with (8), either exactly or to some high order, but the preceding calculations demonstrate that the difference in general is of order $O(n^{-\frac{1}{2}})$, reducing to $O(n^{-1})$ if the parameters are orthogonal.

Acknowledgements

This research was supported in part by National Science Foundation Grant No. DMS 91-01333. This manuscript was prepared using computer facilities supported in part by the National Science Foundation Grants No. DMS 89-05292, DMS 87-03942, DMS 86-01732 and DMS 84-04941 awarded to the Department of Statistics at The University of Chicago, and by the University of Chicago Block Fund.

References

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

BARNDORFF-NIELSEN, O. E. (1988). *Parametric Statistical Models and Likelihood*, Lecture Notes in Statistics 50. Heidelberg: Springer-Verlag.

BARNDORFF-NIELSEN, O. E., BLÆSILD, P. & ERIKSEN, P. S. (1989). *Decomposition and Invariance of Measures, and Statistical Transformation Models*, Lecture Notes in Statistics 58. Heidelberg: Springer-Verlag.

BARNDORFF-NIELSEN, O. E. & JUPP, P. E. (1988). Differential geometry, profile likelihood, L-sufficiency and composite transformation models. *Ann. Statist.* **16**, 1009-43.

COX, D. R. (1975). A note on partially Bayes inference and the linear model. *Biometrika* **62**, 651-4.

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

FISHER, R. A. (1931). The sampling error of estimated deviates, together with other illustrations of the properties and applications of integrals and derivatives of the normal error function. In *Collected Papers of R.A. Fisher*, **2**, Ed. J. H. Bennett, pp. 517-26. University of Adelaide Press.

FISHER, R. A. (1934). Two new properties of mathematical likelihood. *Proc. R. Soc.* A **44**, 285-307.

[*Received January* 1992. *Revised November* 1992]