

Parametrization-invariant Wald tests

P. V. LARSEN¹ and P. E. JUPP²

¹*Department of Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK.
E-mail: p.v.larsen@open.ac.uk*

²*School of Mathematics and Statistics, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK. E-mail: pej@st-andrews.ac.uk*

Parametrization-invariant versions of Wald statistics are introduced, based on yoke geometry. The construction is illustrated by several examples. In the case of simple null hypotheses, formulae are given for generalized Bartlett corrections of these geometric Wald statistics which bring their null distributions close to their large-sample asymptotic distributions.

Keywords: expected geometry; generalized Bartlett correction; observed geometry; parametrization-invariance; Wald statistic; yoke

1. Introduction

One of the key statistics for testing hypotheses in parametric models is the Wald statistic. It has long been known that Wald statistics have the serious drawback that their values depend on the parametrization of the statistical model. As it is customary to compare the value of the Wald statistic with its limiting large-sample null distribution, which does not depend on the parametrization, different parametrizations can easily lead to contradictory conclusions. Parametrization-invariant versions of Wald tests which have been suggested previously include those of Le Cam (1990) and Critchley *et al.* (1996). A considerable disadvantage of these tests is that, in general, computation of the statistics is complicated. For Le Cam's tests, which are based on Hellinger distance, the complexity arises from the need to integrate a product of square roots of probability density functions. For the tests of Critchley *et al.*, which are based on geodesic distance, the complexity arises from the need to integrate a second-order differential equation.

The aim of this paper is to introduce much simpler parametrization-invariant versions of Wald statistics. These new statistics are called 'geometric Wald statistics'. Like the statistics of Critchley *et al.*, the geometric Wald statistics are based on some differential-geometric ideas. The reason why the geometric Wald statistics are simpler than those of Critchley *et al.* is that our construction exploits the linear structure of a vector space (a tangent space to the parameter space), whereas theirs takes place in the parameter space itself. Under the null hypothesis, the large-sample distributions of the geometric Wald statistics are chi-squared. The chi-squared approximation can be improved by generalized Bartlett corrections, which are given in the case of simple null hypotheses.

In Section 2 a geometric interpretation of the Wald statistic is given, and in Section 3 a family of geometric Wald statistics is defined. The geometric Wald statistics are based on

canonical local coordinate systems which are provided automatically by the parametric model under consideration. The construction can be expressed neatly in terms of expected and observed likelihood yokes (Barndorff-Nielsen and Cox 1994, Section 5.6), and can be extended readily to the setting of general yokes. In Section 4 this construction is illustrated by deriving explicit expressions for the geometric Wald statistics for three particular parametric models. In Section 5 generalized Bartlett corrections are given for the geometric Wald statistics in the special case of simple null hypotheses.

We conclude this section 1 by introducing some terminology and notation. Consider a parametric statistical model with probability density function $p(x; \theta)$ with respect to some dominating measure. The parameter θ runs through the parameter space Θ , which is, in general, a manifold but may be considered locally as an open subset of \mathbb{R}^r , so that $\theta = (\theta^1, \dots, \theta^r)$ in some parametrization of Θ . Let ψ be a p -dimensional interest parameter and let χ be a q -dimensional nuisance parameter, such that $\theta = (\psi^1, \dots, \psi^p, \chi^1, \dots, \chi^q)$. Because there is usually no canonical choice of nuisance parameters, only interest-respecting reparametrizations of Θ are considered – that is, under reparametrization of (ψ, χ) to (ϕ, ξ) , the new interest parameter ϕ depends only on ψ and not on χ . The null hypothesis considered is $H_0 : \psi = \psi_0$, which is to be tested against the alternative hypothesis $H_1 : \theta \in \Theta$.

The log-likelihood function based on independent observations x_1, \dots, x_n from the distribution with probability density function $p(x; \theta)$ is denoted by $l(\theta; x_1, \dots, x_n)$. The maximum likelihood estimates of θ under the alternative hypothesis and under the null hypothesis are denoted by $\hat{\theta} = (\hat{\psi}, \hat{\chi})$ and $\tilde{\theta} = (\psi_0, \tilde{\chi})$, respectively. It is required throughout that the Fisher information is defined and that the maximum likelihood estimators are consistent. In Section 5 it is assumed further that the log-likelihood functions are at least four times continuously differentiable with respect to θ , and that all relevant moments of the derivatives of the log-likelihood functions exist.

2. Wald tests

The Wald statistic for testing the simple null hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta \in \Theta$ is

$$W = n(\hat{\theta} - \theta_0)i(\hat{\theta})(\hat{\theta} - \theta_0)^T, \quad (2.1)$$

where $i(\theta)$ denotes the (per-observation) Fisher information matrix, which has elements $i_{i,j}(\theta) = E[-\partial^2 l(\theta; x)/\partial\theta^i\partial\theta^j]$. The two standard generalizations of (2.1) for testing the composite null hypothesis $H_0 : \psi = \psi_0$ against $H_1 : \theta \in \Theta$ are

$$W = n(\hat{\theta} - \tilde{\theta})i(\hat{\theta})(\hat{\theta} - \tilde{\theta})^T \quad (2.2)$$

and

$$W = n(\hat{\psi} - \psi_0)i_{\psi\psi}(\hat{\theta})(\hat{\psi} - \psi_0)^T, \quad (2.3)$$

where $i_{\psi\psi}(\theta)$ denotes the interest part of $i(\theta)$. The intuitive idea of (2.2) is that W is the squared distance between the two maximum likelihood estimates $\hat{\theta}$ and $\tilde{\theta}$, where distance is

measured using the Fisher information metric at $\hat{\theta}$; whereas the intuitive idea of (2.3) is that it is the squared distance between the maximum likelihood estimate $\hat{\psi}$ and the hypothesized value ψ_0 of the interest parameter ψ , where distance is measured using the interest part of the Fisher information metric at $\hat{\theta}$. In general, (2.2) and (2.3) are different. However, they coincide when the parameter space Θ splits as $\Theta = \Psi \times X$ and $\tilde{\chi} = \hat{\chi}$. In particular, this occurs for models with cuts. Cuts are generalizations of sufficient statistics and of ancillary statistics. The definition and examples of cuts can be found in Barndorff-Nielsen and Cox (1994, p. 38) and Lindsey (1996, Section 6.2). For ease of presentation, we shall consider only Wald tests of the form (2.2), unless otherwise specified, although similar results hold for (2.3) throughout. Under H_0 , W is asymptotically χ^2 -distributed with p degrees of freedom, with error of order $O(n^{-1/2})$, in the sense that $P(W > x) = P(\chi_p^2 > x) + O(n^{-1/2})$.

It is well known that, for any given set of data, the value of the Wald statistic W depends on the parametrization used. Indeed, for any given data set, the parametrization can be chosen to give W any positive value. See, for example, Breusch and Schmidt (1988), Gregory and Veall (1985) or Phillips and Park (1988). This lack of parametrization invariance is a considerable disadvantage, because the statistic W is usually compared with its asymptotic χ^2 null distribution, which does not depend on the parametrization. Thus the conclusions drawn from any data set depend on the parametrization of the model. In the next section, we introduce geometric Wald statistics, which are parametrization-invariant analogues of W . As indicated in Remark 3.3, parametrization-invariant analogues of the ‘interest-parameter Wald statistic’ (2.3) can be obtained by using profile likelihood.

3. Geometric Wald tests

The reason for the lack of invariance of the Wald statistic W under reparametrization from θ to η , say, is that, whereas the Fisher information at $\hat{\theta}$ changes in a bilinear way by

$$i(\hat{\eta}) = \frac{\partial \theta}{\partial \eta^T}(\hat{\eta}) i(\hat{\theta}) \frac{\partial \theta^T}{\partial \eta}(\hat{\eta}),$$

the discrepancy between the unrestricted and the restricted maximum likelihood estimates changes from $\hat{\theta} - \tilde{\theta}$ to $\hat{\eta} - \tilde{\eta}$ in a much more complicated way, so that these changes do not ‘cancel out’. A more geometrical description of this is that the Fisher information at $\hat{\theta}$ is a tensor on the tangent space $T_{\hat{\theta}}\Theta$ to the parameter space Θ at $\hat{\theta}$, and so does not depend on the parametrization θ , whereas $\hat{\theta} - \tilde{\theta}$ takes values in a space unrelated to $T_{\hat{\theta}}\Theta$ and does depend on the parametrization.

The key to obtaining the parametrization-invariant versions of Wald statistics considered here is to measure the discrepancy between the unrestricted maximum likelihood estimate $\hat{\theta}$ and the restricted maximum likelihood estimate $\tilde{\theta}$ by some vector $\Gamma_{\hat{\theta}}(\tilde{\theta})$ which changes linearly under reparametrization from θ to η by

$$\Gamma_{\hat{\eta}}(\tilde{\eta}) = \Gamma_{\hat{\theta}}(\tilde{\theta}) \frac{\partial \eta}{\partial \theta^T}(\hat{\theta}). \tag{3.1}$$

Then, under reparametrization, the changes in (3.1) cancel with the changes in the matrix

expression for the Fisher information to yield the parametrization-invariant scalar $\Gamma_{\hat{\theta}}(\tilde{\theta})i(\hat{\theta})\Gamma_{\hat{\theta}}(\tilde{\theta})^T$. The $\Gamma_{\hat{\theta}}$ can be regarded as a local coordinate system which takes values in the tangent space $T_{\hat{\theta}}\Theta$ to Θ at $\hat{\theta}$. We now show how suitable $\Gamma_{\hat{\theta}}$ are given naturally by the model itself.

A standard way of measuring the ‘distance’ between two points θ and θ' in the parameter space of a parametric statistical model is by the Kullback–Leibler divergence

$$K(\theta, \theta') = E_{\theta}\{l(\theta; x) - l(\theta'; x)\}.$$

For our purposes, it is useful to consider instead the expected likelihood yoke of the model, which is the function f on $\Theta \times \Theta$ given by

$$f(\theta; \theta') = E_{\theta'}\{l(\theta; x) - l(\theta'; x)\}, \quad (3.2)$$

so that $f(\theta; \theta') = -K(\theta', \theta)$. It is straightforward to verify that the functions $\overset{1}{\Gamma}_{\theta}$ defined by

$$\overset{1}{\Gamma}_{\theta}(\theta') = \frac{\partial f}{\partial \theta}(\theta'; \theta)i(\theta)^{-1} \quad (3.3)$$

satisfy (3.1). Since $\overset{1}{\Gamma}_{\theta}$ has non-singular derivative at θ , it is a local coordinate system in some neighbourhood of θ . The Kullback–Leibler divergence is not symmetrical in its arguments, and we could just as well differentiate it with respect to the first argument instead of the second. This yields another set of functions $\overset{-1}{\Gamma}_{\theta}$ given by

$$\overset{-1}{\Gamma}_{\theta}(\theta') = \frac{\partial f}{\partial \theta'}(\theta; \theta')i(\theta)^{-1}, \quad (3.4)$$

which satisfy (3.1) and form local coordinate systems around θ . Taking appropriate linear combinations of (3.3) and (3.4) yields the one-parameter family $\overset{\alpha}{\Gamma}_{\theta}$ of local coordinate systems on Θ around θ , defined by

$$\overset{\alpha}{\Gamma}_{\theta}(\theta') = \frac{1 + \alpha}{2} \overset{1}{\Gamma}_{\theta}(\theta') + \frac{1 - \alpha}{2} \overset{-1}{\Gamma}_{\theta}(\theta'), \quad (3.5)$$

for any real α .

Note that the per-observation Fisher information $i(\theta)$ satisfies

$$i(\theta) = \frac{\partial^2}{\partial \theta \partial \theta'^T} f(\theta; \theta')|_{\theta'=\theta}.$$

Thus, the statistics $\Gamma_{\hat{\theta}}(\tilde{\theta})i(\hat{\theta})\Gamma_{\hat{\theta}}(\tilde{\theta})^T$ are defined entirely in terms of the expected likelihood yoke (3.2). The appropriate setting for these statistics is that of general yokes. A yoke on a parameter space Θ is a real-valued function g on $\Theta \times \Theta$ such that

$$\frac{\partial}{\partial \theta} g(\theta; \theta')|_{\theta'=\theta} = 0 \quad (3.6)$$

and

$$\gamma(\theta) = \frac{\partial^2}{\partial \theta \partial \theta'^T} g(\theta; \theta')|_{\theta'=\theta} \text{ is non-singular,} \quad (3.7)$$

for all θ in Θ . The geometrical interpretation of the second mixed derivative γ is that it is the

semi-Riemannian metric determined by g . See Barndorff-Nielsen and Cox (1994, Section 5.6) and Barndorff-Nielsen *et al.* (1994, Section 3.3). A yoke g on Θ is called normalized if

$$g(\theta; \theta) = 0 \quad (3.8)$$

for all θ in Θ .

Remark 3.1. Every yoke gives rise to a preferred point geometry (in a slightly weaker sense than that of Critchley *et al.*, 1993; 1994). The precise relationship between yokes and preferred point geometries can be found in Remark 3.1 of Barndorff-Nielsen and Jupp (1997).

For any yoke g , local coordinate systems $\overset{1}{\Gamma}_\theta$ which satisfy (3.1) can be defined by

$$\overset{1}{\Gamma}_\theta(\theta') = \frac{\partial g}{\partial \theta}(\theta'; \theta)\gamma(\theta)^{-1}, \quad (3.9)$$

where γ is given by (3.7). Differentiating with respect to the first argument instead of the second yields the local coordinate system $\overset{-1}{\Gamma}_\theta$ with

$$\overset{-1}{\Gamma}_\theta(\theta') = \frac{\partial g}{\partial \theta'}(\theta; \theta')\gamma(\theta)^{-1}. \quad (3.10)$$

The coordinate systems $\overset{1}{\Gamma}_\theta$ and $\overset{-1}{\Gamma}_\theta$ arise naturally from the yoke g . They are special cases of a one-parameter family $\overset{\alpha}{\Gamma}_\theta$ of local coordinate systems on Θ around θ , defined by

$$\overset{\alpha}{\Gamma}_\theta(\theta') = \frac{1 + \alpha}{2} \overset{1}{\Gamma}_\theta(\theta') + \frac{1 - \alpha}{2} \overset{-1}{\Gamma}_\theta(\theta'), \quad (3.11)$$

for any real α . If g is the expected likelihood yoke (3.2) then (3.9)–(3.11) become (3.3)–(3.5).

It follows from (3.6) and (3.8) that, for any normalized yoke and for all α ,

$$\overset{\alpha}{\Gamma}_\theta(\theta) = 0, \quad (3.12)$$

so that, in the local coordinate system $\overset{\alpha}{\Gamma}$ around θ , θ behaves as the origin. Further details of the $\overset{\alpha}{\Gamma}_\theta$ coordinate systems can be found in Barndorff-Nielsen and Cox (1994, Section 5.6) and Barndorff-Nielsen *et al.* (1994, Section 3.3). Geometrically minded readers may wish to note that, in contrast to more familiar coordinate systems which map Θ to a fixed vector space, $\overset{\alpha}{\Gamma}_\theta$ maps Θ to the tangent space $T_\theta\Theta$ at θ , and this is a vector space which depends on θ .

Two natural choices of α are $\alpha = 1$ and $\alpha = -1$, which give rise to the local coordinate systems $\overset{1}{\Gamma}_\theta$ and $\overset{-1}{\Gamma}_\theta$, respectively. Any value of α between 1 and -1 corresponds to a weighted average of these two local coordinate systems, with the special case $\alpha = 0$ giving equal weight to both. It is observed in Section 5 that the values $\alpha = \pm\frac{1}{3}$ have some nice properties, as do $\alpha = \pm 1$.

Remark 3.2. There is no relationship between the $\overset{\alpha}{\Gamma}_\theta$ local coordinate systems and the

parametrizations of one-parameter curved exponential models considered by Hougaard (1982) and Kass (1984). Note that the $\overset{\alpha}{\Gamma}_{\theta}$ exist for all parametric models and depend on the point θ .

The fact that the $\overset{\alpha}{\Gamma}_{\tilde{\theta}}(\tilde{\theta})$ in (3.11) obey (3.1) suggests the following general definitions. Let g be any yoke on Θ . Then, for any real α , the corresponding geometric Wald statistic $\overset{\alpha}{\mathcal{W}}$ is

$$\overset{\alpha}{\mathcal{W}} = n \overset{\alpha}{\Gamma}_{\tilde{\theta}}(\tilde{\theta}) \gamma(\tilde{\theta}) \overset{\alpha}{\Gamma}_{\tilde{\theta}}(\tilde{\theta})^T \tag{3.13}$$

where γ is given in (3.7). The statistic $\overset{\alpha}{\mathcal{W}}$ is parametrization-invariant, and it follows from (3.12) that if $\tilde{\theta} = \hat{\theta}$ then $\overset{\alpha}{\mathcal{W}} = 0$ for all α . Equations (3.12) and (3.13) show that $\overset{\alpha}{\mathcal{W}}$ can be regarded as using the metric γ to measure a squared distance between $\tilde{\theta}$ and $\hat{\theta}$. In this respect, it retains the idea of the traditional Wald statistic (2.2). Note that equation (3.13) can be considered as a quadratic approximation to the likelihood ratio statistic, $w = 2 \{l(\hat{\theta}; x) - l(\tilde{\theta}; x)\}$, in the $\overset{\alpha}{\Gamma}_{\tilde{\theta}}$ coordinate system.

Apart from the expected likelihood yokes, the main yokes of interest in parametric inference are the observed likelihood yokes. They, too, arise naturally from the model. An observed likelihood yoke requires an auxiliary statistic a such that $(\hat{\theta}, a)$ is minimal sufficient for θ . For each fixed value of a , the observed likelihood yoke g is given by

$$\begin{aligned} g(\theta; \theta') &= n^{-1} \{l(\theta; x_1, \dots, x_n) - l(\theta'; x_1, \dots, x_n)\} \\ &= n^{-1} \{l(\theta; \theta', a) - l(\theta'; \theta', a)\}, \end{aligned} \tag{3.14}$$

where (x_1, \dots, x_n) satisfies $\hat{\theta}(x_1, \dots, x_n) = \theta'$. The observed likelihood yoke is usually defined as $l(\theta; x_1, \dots, x_n) - l(\theta'; x_1, \dots, x_n)$, but it is convenient to use definition (3.14) here, in order to make $g(\theta; \theta')$ of order $O(1)$ for any fixed (θ, θ') . Note that the derivative $\partial g(\theta; \hat{\theta}, a) / \partial \theta$ is the per-observation score vector, and the metric γ obtained from the observed likelihood yoke is the per-observation observed information

$$j(\theta) = \frac{\partial^2}{\partial \theta \partial \theta^T} g(\theta; \theta')|_{\theta'=\theta}. \tag{3.15}$$

When it is necessary to emphasize which likelihood yoke is employed in constructing the geometric Wald statistics, subscripts will be used, so that, for example, $\overset{\alpha}{\mathcal{W}}_e$ and $\overset{\alpha}{\mathcal{W}}_o$ denote geometric Wald statistics based on the expected and observed likelihood yokes, respectively. If the auxiliary statistic a is ancillary, or approximately ancillary to the extent specified in Barndorff-Nielsen and Cox (1994, pp. 176–177), then standard results (Barndorff-Nielsen and Cox, 1994, p. 158) on the closeness of observed and expected likelihood yokes show that W , $\overset{\alpha}{\mathcal{W}}_e$ and $\overset{\alpha}{\mathcal{W}}_o$ are equal to order $O_p(n^{-1/2})$.

Remark 3.3. The geometric Wald statistics defined in (3.13) are like the Wald statistic (2.2) in being based on the full parameter θ . Geometric Wald statistics which are analogous to the ‘interest-parameter Wald statistic’ (2.3) in being based only on the interest parameter ψ can be constructed from any suitable pseudo-likelihood on the space Ψ of interest parameters. The construction uses the yoke on Ψ given by the pseudo-likelihood. For models with cuts, the geometric Wald statistics constructed using the profile likelihood yoke coincide with

those constructed using the observed likelihood yoke. The geometric Wald statistic $\overset{1}{\mathcal{W}}_o$ constructed using the profile likelihood yoke occurs in the expression in Barndorff-Nielsen and Cox (1994, equation (6.140)) for the part z_{INF} of the Pierce and Peters (1992) decomposition of the modified directed likelihood r^* .

A variant of the Wald statistic W is the modified Wald statistic \tilde{W} of Hayakawa and Puri (1985), which is obtained from (2.2) by interchanging $\hat{\theta}$ and $\tilde{\theta}$, that is,

$$\tilde{W} = n(\hat{\theta} - \tilde{\theta})i(\tilde{\theta})(\hat{\theta} - \tilde{\theta})^T. \tag{3.16}$$

The value of \tilde{W} (like the value of W) depends on the parametrization. Invariant versions of this modified test statistic can be provided by interchanging $\hat{\theta}$ and $\tilde{\theta}$ in (3.13), resulting in the modified geometric Wald statistic

$$\overset{\alpha}{\tilde{\mathcal{W}}} = n \overset{\alpha}{\Gamma}_{\tilde{\theta}}(\hat{\theta}) \gamma(\tilde{\theta}) \overset{\alpha}{\Gamma}_{\hat{\theta}}(\tilde{\theta})^T. \tag{3.17}$$

Note that (3.17) can be considered as a quadratic approximation to the likelihood ratio statistic, $w = 2\{l(\hat{\theta}; x) - l(\tilde{\theta}; x)\}$, in the $\overset{\alpha}{\Gamma}_{\tilde{\theta}}$ coordinate system. First-order Taylor expansions show that the traditional Wald statistics W and \tilde{W} , and the geometric Wald statistics $\overset{\alpha}{\mathcal{W}}_e$ and $\overset{\alpha}{\tilde{\mathcal{W}}}_e$, are all equal to order $O_P(n^{-1/2})$. If the auxiliary statistic a is ancillary, or approximately ancillary, then W , \tilde{W} , $\overset{\alpha}{\mathcal{W}}_e$, $\overset{\alpha}{\tilde{\mathcal{W}}}_e$, $\overset{\alpha}{\mathcal{W}}_o$ and $\overset{\alpha}{\tilde{\mathcal{W}}}_o$ are all equal to order $O_P(n^{-1/2})$.

An important special case of the modified geometric Wald tests occurs when the yoke g is the observed likelihood yoke (3.14). Then

$$\overset{-1}{\mathcal{W}}_o = n \left\{ \frac{1}{n} \frac{\partial l}{\partial \theta}(\tilde{\theta}) \right\} j(\tilde{\theta})^{-1} \left\{ \frac{1}{n} \frac{\partial l}{\partial \theta^T}(\tilde{\theta}) \right\}, \tag{3.18}$$

which can be regarded as a variant of the quadratic score statistic

$$S = \frac{\partial l}{\partial \theta}(\tilde{\theta}) \{ ni(\tilde{\theta}) \}^{-1} \frac{\partial l}{\partial \theta^T}(\tilde{\theta}), \tag{3.19}$$

obtained by using the observed information instead of the Fisher information. Thus, in models for which the observed information j is equal to the expected information i , $\overset{-1}{\mathcal{W}}_o = S$. In particular, this is true for full exponential models. Moreover, for multivariate normal distributions with known variance, it is straightforward to see that, for either likelihood yoke and for any α , $\overset{\alpha}{\mathcal{W}} = \overset{\alpha}{\tilde{\mathcal{W}}} = S = w$. Note that, in general, $\overset{-1}{\mathcal{W}}_e$ can be regarded as an ‘expected’ analogue of S . Thus the class of geometric Wald statistics includes statistics in the spirit of both Wald statistics and quadratic score statistics.

In general, $\overset{\alpha}{\mathcal{W}}$ and $\overset{\alpha}{\tilde{\mathcal{W}}}$ are different and depend on α . There are two reasons why the choice of α is not crucial. Firstly, if the yoke g is symmetrical, in that $g(\theta, \theta') = g(\theta', \theta)$, then neither $\overset{\alpha}{\mathcal{W}}$ nor $\overset{\alpha}{\tilde{\mathcal{W}}}$ depends on α . Secondly, the generalized Bartlett corrections $\overset{\alpha}{\mathcal{W}}^*$ and $\overset{\alpha}{\tilde{\mathcal{W}}}^*$ of $\overset{\alpha}{\mathcal{W}}$ and $\overset{\alpha}{\tilde{\mathcal{W}}}$ introduced in Section 5, have distributions which depend on α only to order $O(n^{-2})$. Since neither $\overset{\alpha}{\mathcal{W}}$ nor its modified version $\overset{\alpha}{\tilde{\mathcal{W}}}$ seems to have a general distributional advantage over the other, it is sensible to use the one which is easier to compute in any given problem.

4. Examples

In this section three examples are used to illustrate the construction and properties of the geometric Wald statistics $\overset{\alpha}{\mathcal{W}}$ and $\overset{1}{\mathcal{W}}$.

Example 1. Full exponential models. Consider a full exponential model with density function

$$p(x; \theta) = \exp\{\theta t(x)^T - \kappa(\theta)\}, \quad (4.1)$$

where t is the canonical statistic and θ is the canonical parameter. Let η be the expectation parameter, that is, $\eta = \eta(\theta) = E_{\theta}\{t(X)\}$, and let $i(\theta)$ denote the Fisher information matrix in the θ -parametrization. Since $\eta = \partial\kappa(\theta)/\partial\theta$, the Fisher information matrix in the η -parametrization is $i(\eta) = i(\theta)^{-1}$. The two likelihood yokes coincide and are equal to

$$g(\theta; \theta') = n^{-1} \left\{ (\theta - \theta') \frac{\partial\kappa}{\partial\theta}(\theta') - \kappa(\theta) + \kappa(\theta') \right\}.$$

Then the special local coordinate systems given by (3.9) and (3.10) take the forms

$$\overset{1}{\Gamma}_{\theta}(\theta') = n^{-1}(\theta' - \theta)$$

and

$$\overset{-1}{\Gamma}_{\theta}(\theta') = n^{-1}(\eta' - \eta)i(\theta)^{-1},$$

so that these coordinate systems are affine functions of the canonical parametrization and the expectation parametrization, respectively, of the exponential model. Thus, for full exponential models, the local coordinate systems $\overset{1}{\Gamma}_{\theta}$ and $\overset{-1}{\Gamma}_{\theta}$ are actually global coordinate systems. Calculation shows that the geometric Wald statistics are

$$\overset{\alpha}{\mathcal{W}} = n \left\{ \frac{1+\alpha}{2}(\hat{\theta} - \tilde{\theta})i(\hat{\theta}) + \frac{1-\alpha}{2}(\hat{\eta} - \tilde{\eta}) \right\} i(\hat{\theta})^{-1} \left\{ \frac{1+\alpha}{2}(\hat{\theta} - \tilde{\theta})i(\hat{\theta}) + \frac{1-\alpha}{2}(\hat{\eta} - \tilde{\eta}) \right\}^T,$$

and the modified geometric Wald statistics are

$$\overset{\alpha}{\tilde{\mathcal{W}}} = n \left\{ \frac{1+\alpha}{2}(\tilde{\theta} - \hat{\theta})i(\tilde{\theta}) + \frac{1-\alpha}{2}(\tilde{\eta} - \hat{\eta}) \right\} i(\tilde{\theta})^{-1} \left\{ \frac{1+\alpha}{2}(\tilde{\theta} - \hat{\theta})i(\tilde{\theta}) + \frac{1-\alpha}{2}(\tilde{\eta} - \hat{\eta}) \right\}^T,$$

In particular,

$$\overset{1}{\mathcal{W}} = n(\hat{\theta} - \tilde{\theta})i(\hat{\theta})(\hat{\theta} - \tilde{\theta})^T \quad \text{and} \quad \overset{-1}{\mathcal{W}} = n(\hat{\eta} - \tilde{\eta})i(\hat{\eta})(\hat{\eta} - \tilde{\eta})^T$$

are the traditional Wald statistics (2.2) using the canonical and expectation parametrization, respectively. Furthermore,

$${}^1\mathcal{W} = n(\hat{\theta} - \tilde{\theta})i(\tilde{\theta})(\hat{\theta} - \tilde{\theta})^T \quad \text{and} \quad \tilde{\mathcal{W}} = S = n(\hat{\eta} - \tilde{\eta})i(\tilde{\eta})(\hat{\eta} - \tilde{\eta})^T$$

are the modified traditional Wald statistics (3.16) in the canonical and expectation parametrization, respectively.

Example 2. Simple linear regression. Let Y_1, \dots, Y_n be independent normal random variables with unknown variances σ^2 and means $E(Y_i) = \alpha + \beta(x_i - \bar{x})$, respectively, where x_1, \dots, x_n are known constants. The expected and observed likelihood yokes coincide and are given by

$$f(\theta; \theta') = \frac{1}{2} \left(\log \sigma'^2 - \log \sigma^2 + 1 - \frac{\sigma'^2}{\sigma^2} \right) - \frac{1}{2\sigma^2} \{(\alpha' - \alpha)^2 + (\beta' - \beta)^2 \bar{S}_{xx}\},$$

where $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ and $\bar{S}_{xx} = n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$. Calculation shows that the geometric Wald statistics (3.13) are

$$\begin{aligned} {}^\gamma\mathcal{W} &= n\hat{\sigma}^2 \left(\frac{1+\gamma}{2\tilde{\sigma}^2} + \frac{1-\gamma}{2\hat{\sigma}^2} \right)^2 \{(\tilde{\alpha} - \hat{\alpha})^2 + (\tilde{\beta} - \hat{\beta})^2 \bar{S}_{xx}\} \\ &\quad + \frac{n}{2} \left[\frac{\hat{\sigma}^4 - \tilde{\sigma}^4 - \gamma(\tilde{\sigma}^2 - \hat{\sigma}^2)^2}{2\tilde{\sigma}^2\hat{\sigma}^2} + \frac{1-\gamma}{2\hat{\sigma}^2} \{(\tilde{\alpha} - \hat{\alpha})^2 + (\tilde{\beta} - \hat{\beta})^2 \bar{S}_{xx}\} \right]^2. \end{aligned}$$

The modified geometric Wald statistics (3.17) follow by interchanging $\hat{\theta}$ and $\tilde{\theta}$ in the above formula, so that

$$\begin{aligned} \tilde{{}^\gamma\mathcal{W}} &= n\tilde{\sigma}^2 \left(\frac{1+\gamma}{2\hat{\sigma}^2} + \frac{1-\gamma}{2\tilde{\sigma}^2} \right)^2 \{(\tilde{\alpha} - \hat{\alpha})^2 + (\tilde{\beta} - \hat{\beta})^2 \bar{S}_{xx}\} \\ &\quad + \frac{n}{2} \left[\frac{\tilde{\sigma}^4 - \hat{\sigma}^4 - \gamma(\tilde{\sigma}^2 - \hat{\sigma}^2)^2}{2\tilde{\sigma}^2\hat{\sigma}^2} + \frac{1-\gamma}{2\tilde{\sigma}^2} \{(\tilde{\alpha} - \hat{\alpha})^2 + (\tilde{\beta} - \hat{\beta})^2 \bar{S}_{xx}\} \right]^2. \end{aligned}$$

When the null hypothesis is $H_0 : \beta = 0$, the geometric Wald statistics and modified geometric Wald statistics reduce to

$${}^\gamma\mathcal{W} = n \left[\frac{r^2 \{1 - (1 + \gamma)r^2/2\}^2}{1 - r^2} + \frac{r^4 \{3 - \gamma - (1 + \gamma)r^2\}^2}{8(1 - r^2)^2} \right]$$

and

$$\tilde{{}^\gamma\mathcal{W}} = n \left[\frac{r^2 \{1 - (1 - \gamma)r^2/2\}^2}{(1 - r^2)^2} + \frac{r^4 (1 + \gamma)^2}{8(1 - r^2)^2} \right],$$

where r denotes the sample correlation coefficient. In four important cases these simplify to

$$\begin{aligned} \mathcal{W} &= nr^2 \left(1 - \frac{r^2}{2}\right), & \mathcal{W}^{-1} &= \frac{nr^2(1+r^2)}{(1-r^2)^2}, \\ \tilde{\mathcal{W}} &= \frac{nr^2(1+r^2/2)}{(1-r^2)^2}, & \tilde{\mathcal{W}}^{-1} &= S = nr^2. \end{aligned}$$

The final example concerns a simple model in which the expected and observed likelihood yokes differ.

Example 3. Fisher's gamma hyperbola. Let X_i and Y_i ($i = 1, \dots, n$) be independent exponentially distributed random variables with $E(X_i) = 1/\theta$ and $E(Y_i) = \theta$, $\theta > 0$ (Barndorff-Nielsen and Cox 1994, p. 193). Then the log-likelihood function is

$$l(\theta; x_1, \dots, x_n, y_1, \dots, y_n) = -n \left(\theta \bar{x} + \frac{1}{\theta} \bar{y} \right),$$

where $\bar{x} = n^{-1} \sum_{i=1}^n x_i$ and $\bar{y} = n^{-1} \sum_{i=1}^n y_i$. The log-likelihood function can be expressed in terms of $\hat{\theta} = (\bar{y}/\bar{x})^{1/2}$ and the ancillary $a = (\bar{x}\bar{y})^{1/2}$ by

$$l(\theta; \hat{\theta}, a) = -na \left(\frac{\theta}{\hat{\theta}} + \frac{\hat{\theta}}{\theta} \right).$$

Then the expected likelihood yoke (3.2) is

$$f(\theta; \theta') = 2 - \frac{\theta}{\theta'} - \frac{\theta'}{\theta},$$

and the observed likelihood yoke (3.14) is

$$\begin{aligned} g(\theta; \theta') &= a \left(2 - \frac{\theta}{\theta'} - \frac{\theta'}{\theta} \right) \\ &= af(\theta; \theta'). \end{aligned}$$

Both likelihood yokes are symmetric in θ and θ' , so that the Γ_θ^α of (3.11) do not depend on the value of α . Thus, in either geometry, the geometric Wald statistics \mathcal{W}^α and $\tilde{\mathcal{W}}^\alpha$ are equal and do not depend on α . Straightforward calculations show that the geometric Wald statistics based on expected geometry are

$$\mathcal{W}_e = \tilde{\mathcal{W}}_e = \frac{n}{2} \left(\frac{\tilde{\theta}^2 - \hat{\theta}^2}{\hat{\theta}\tilde{\theta}} \right)^2 \quad (4.2)$$

and the geometric Wald statistics based on observed geometry are

$$\mathcal{W}_o = \tilde{\mathcal{W}}_o = \frac{n}{2} a \left(\frac{\tilde{\theta}^2 - \hat{\theta}^2}{\hat{\theta}\tilde{\theta}} \right)^2. \quad (4.3)$$

In comparison, the score statistic (3.19) is

$$S = \frac{n}{2} a^2 \left(\frac{\tilde{\theta}^2 - \hat{\theta}^2}{\hat{\theta}\tilde{\theta}} \right)^2,$$

so that $S = a\mathcal{W}_o = a^2\mathcal{W}_e$. Thus changes in the ancillary a have a greater effect on the score statistic than on the geometric Wald statistics. Since (\bar{x}, \bar{y}) tends almost surely to $(1/\theta, \theta)$, the ratios $\mathcal{W}_o/\mathcal{W}_e$ and S/\mathcal{W}_o tend to 1 almost surely as n tends to infinity.

5. Generalized Bartlett correction

For any statistic S having a χ_p^2 null distribution with error of order $O(n^{-1/2})$, that is, $P(S > x) = P(\chi_p^2 > x) + O(n^{-1/2})$ under the null hypothesis, Cordeiro and Ferrari (1991) showed that there is a polynomial modification S^* of S satisfying

$$P(S^* > x) = P(\chi_p^2 > x) + O(n^{-3/2}).$$

They also showed that when S is the score statistic, S^* is a cubic function of S , of the form

$$S^* = \left\{ 1 - \frac{1}{n}(c + bS + aS^2) \right\} S, \quad (5.1)$$

for suitable coefficients a , b and c . An argument similar to that of Barndorff-Nielsen and Hall (1988) shows that (5.1) has error of order $O(n^{-2})$ rather than $O(n^{-3/2})$, that is,

$$P(S^* > x) = P(\chi_p^2 > x) + O(n^{-2}). \quad (5.2)$$

The method of Cordeiro and Ferrari (1991) and the argument of Barndorff-Nielsen and Hall (1988) can be extended to give generalized Bartlett corrections \mathcal{W}_a^* and $\tilde{\mathcal{W}}_a^*$ of the geometric Wald statistics \mathcal{W}_a and $\tilde{\mathcal{W}}_a$ of the form

$$\mathcal{W}_a^* = \left\{ 1 - \frac{1}{n}(c + b\mathcal{W}_a + a\mathcal{W}_a^2) \right\} \mathcal{W}_a, \quad (5.3)$$

$$\tilde{\mathcal{W}}_a^* = \left\{ 1 - \frac{1}{n}(\tilde{c} + \tilde{b}\tilde{\mathcal{W}}_a + \tilde{a}\tilde{\mathcal{W}}_a^2) \right\} \tilde{\mathcal{W}}_a, \quad (5.4)$$

and with errors of order $O(n^{-2})$.

For simplicity, the coefficients a , b , c and \tilde{a} , \tilde{b} , \tilde{c} are given here only in the case of simple null hypotheses. It is intended that the general results will be presented elsewhere.

To express the coefficients a , b and c concisely, it is convenient to use some tensors derived from the yoke g ; see Blæsild (1991). These tensors are defined by

$$T_{ijk}(\theta) = g_{i,jk}(\theta; \theta) - g_{jk,i}(\theta; \theta) \quad (5.5)$$

$$= -g_{ij;k}(\theta; \theta)[3]_{ijk} - g_{ijk}(\theta; \theta),$$

$$T_{ij,kl}(\theta) = g_{ij,kl}(\theta; \theta) - g_{ij,m}(\theta; \theta)g^{m;n}(\theta)g_{n,kl}(\theta; \theta), \quad (5.6)$$

$$T_{i,jkl}(\theta) = g_{i,jkl}(\theta; \theta) - g_{jkl,i}(\theta; \theta) - T_{ijm}(\theta)g^{m;n}(\theta)g_{kl,n}(\theta; \theta)[3]_{jkl}, \quad (5.7)$$

$$T_{ijkl}(\theta) = -T_{i,jkl}(\theta) - T_{ij,kl}(\theta)[3]_{jkl}, \quad (5.8)$$

where $g^{m;n}$ denotes the (m, n) th element of the inverse of the matrix of second mixed derivatives of g , and the subscripts of g denote the mixed derivatives, for example,

$$g_{i,jk}(\theta_1; \theta_2) = \frac{\partial^3 g}{\partial \theta_1^i \partial \theta_2^j \partial \theta_2^k}(\theta_1; \theta_2),$$

$$g_{ij,kl}(\theta_1; \theta_2) = \frac{\partial^4 g}{\partial \theta_1^i \partial \theta_1^j \partial \theta_2^k \partial \theta_2^l}(\theta_1; \theta_2).$$

In (5.5)–(5.8), the Einstein summation convention has been used and the notation $[3]_{ijk}$ indicates a sum over the indicated number of terms obtained by permuting the subscripts. The tensors (5.5)–(5.8) have proved invaluable in concise invariant expressions (Blæsild 1991; Barndorff-Nielsen and Cox 1994, Section 5.3) for Bartlett corrections of likelihood ratio statistics.

Because the generalized Bartlett corrections involving expected likelihood yokes are based on the unconditional asymptotic distribution of the score (Barndorff-Nielsen and Cox 1989, Section 6.3), whereas the corrections involving observed likelihood yokes are based on the conditional asymptotic distribution of the score given an ancillary statistic (Mora, 1992), the formulae for the coefficients take slightly different forms in the two cases.

In the case of a simple null hypothesis, $H_0 : \theta = \theta_0$, the generalized Bartlett correction for the geometric Wald statistics based on the expected likelihood yoke (3.2) can be written in terms of the tensors (5.5)–(5.8) obtained from the expected likelihood yoke, and the tensors

$$\tau_{i,j,k,l}(\theta) = E_{\theta} \left[\frac{\partial l}{\partial \theta^i}(\theta; x) \frac{\partial l}{\partial \theta^j}(\theta; x) \frac{\partial l}{\partial \theta^k}(\theta; x) \frac{\partial l}{\partial \theta^l}(\theta; x) \right] - i(\theta)_{i,j} i(\theta)_{k,l} [3]_{ijk}$$

$$\begin{aligned} \tau_{ij,kl}(\theta) &= E_{\theta} \left\{ \frac{\partial^2 l}{\partial \theta^i \partial \theta^j}(\theta; x) \frac{\partial l}{\partial \theta^k \partial \theta^l}(\theta; x) \right\} \\ &\quad - E_{\theta} \left[\frac{\partial^2 l}{\partial \theta^i \partial \theta^j}(\theta; x) \frac{\partial l}{\partial \theta^m}(\theta; x) \right] i(\theta)^{m,n} E_{\theta} \left\{ \frac{\partial l}{\partial \theta^n}(\theta; x) \frac{\partial^2 l}{\partial \theta^k \partial \theta^l}(\theta; x) \right\} + i(\theta)_{i,j} i(\theta)_{k,l}. \end{aligned}$$

The coefficients a , b and c in (5.3) obtained from the expected likelihood yoke are given by

$$\begin{aligned}
 a &= \frac{(1-3\alpha)^2}{48p(p+2)(p+4)} i(\theta_0)^{i,j} i(\theta_0)^{k,l} \\
 &\quad \times 3\{T_{ijm}(\theta_0) i(\theta_0)^{m,n} T_{nkl}(\theta_0) + 2T_{ikm}(\theta_0) i(\theta_0)^{m,n} T_{njl}(\theta_0)\}, \\
 b &= \frac{1}{12p(p+2)} i(\theta_0)^{i,j} i(\theta_0)^{k,l} \{3\tau_{i,j,k,l}(\theta_0) - 6(1-2\alpha)T_{i,jkl}(\theta_0) \\
 &\quad + 3(1-2\alpha-2\alpha^2)T_{ijm}(\theta_0) i(\theta_0)^{m,n} T_{nkl}(\theta_0) \\
 &\quad + (5-12\alpha-3\alpha^2)T_{ikm}(\theta_0) i(\theta_0)^{m,n} T_{njl}(\theta_0)\}, \\
 c &= \frac{1}{12p} i(\theta_0)^{i,j} i(\theta_0)^{k,l} \{-3\tau_{i,j,k,l}(\theta_0) - 12T_{ij,kl}(\theta_0) + 12\tau_{ik,jl}(\theta_0) \\
 &\quad + 3T_{ijm}(\theta_0) i(\theta_0)^{m,n} T_{nkl}(\theta_0) + 2T_{ikm}(\theta_0) i(\theta_0)^{m,n} T_{njl}(\theta_0)\},
 \end{aligned}$$

and the coefficients \tilde{a} , \tilde{b} and \tilde{c} in (5.4) obtained from the expected likelihood yoke are given by

$$\begin{aligned}
 \tilde{a} &= \frac{(1+3\alpha)^2}{48p(p+2)(p+4)} i(\theta_0)^{i,j} i(\theta_0)^{k,l} \\
 &\quad \times \{3T_{ijm}(\theta_0) i(\theta_0)^{m,n} T_{nkl}(\theta_0) + 2T_{ikm}(\theta_0) i(\theta_0)^{m,n} T_{njl}(\theta_0)\}, \\
 \tilde{b} &= \frac{1}{12p(p+2)} i(\theta_0)^{i,j} i(\theta_0)^{k,l} \{3\tau_{i,j,k,l}(\theta_0) - 6(1+\alpha)T_{i,jkl}(\theta_0) \\
 &\quad + 3(1-2\alpha-2\alpha^2)T_{ijm}(\theta_0) i(\theta_0)^{m,n} T_{nkl}(\theta_0) \\
 &\quad + (5-12\alpha-3\alpha^2)T_{ikm}(\theta_0) i(\theta_0)^{m,n} T_{njl}(\theta_0)\}, \\
 \tilde{c} &= c.
 \end{aligned}$$

For the geometric Wald statistics based on the observed likelihood yoke, the coefficients a , b and c in (5.3) and \tilde{a} , \tilde{b} and \tilde{c} in (5.4) depend only on the tensors (5.5)–(5.8) obtained from the observed likelihood yoke (3.14). Here, it is required that the auxiliary statistic in the observed yoke (3.14) is ancillary, at least approximately. The coefficients in (5.3) constructed from the observed likelihood yoke are given by

$$\begin{aligned}
a &= \frac{(1 - 3\alpha)^2}{48p(p+2)(p+4)} j(\theta_0)^{i,j} j(\theta_0)^{k,l} \\
&\quad \times \{3T_{ijm}(\theta_0)j(\theta_0)^{m,n}T_{nkl}(\theta_0) + 2T_{ikm}(\theta_0)j(\theta_0)^{m,n}T_{nij}(\theta_0)\}, \\
b &= \frac{1}{12p(p+2)} j(\theta_0)^{i,j} j(\theta_0)^{k,l} \{3T_{ijkl}(\theta_0) + 12\alpha T_{i,jkl}(\theta_0) \\
&\quad + 3(1 - 2\alpha - 2\alpha^2)T_{ijm}(\theta_0)j(\theta_0)^{m,n}T_{nkl}(\theta_0) \\
&\quad + (5 - 12\alpha - 3\alpha^2)T_{ikm}(\theta_0)j(\theta_0)^{m,n}T_{nij}(\theta_0)\}, \\
c &= \frac{1}{12p} j(\theta_0)^{i,j} j(\theta_0)^{k,l} \{3T_{ijkl}(\theta_0) + 12T_{ik,jl}(\theta_0) + 3T_{ijm}(\theta_0)j(\theta_0)^{m,n}T_{nkl}(\theta_0) \\
&\quad + 2T_{ikm}(\theta_0)j(\theta_0)^{m,n}T_{nij}(\theta_0)\}.
\end{aligned}$$

Similarly, the coefficients \tilde{a} , \tilde{b} and \tilde{c} for the modified geometric Wald statistic (5.4) based on the observed likelihood yoke are

$$\begin{aligned}
\tilde{a} &= \frac{(1 + 3\alpha)^2}{48p(p+2)(p+4)} j(\theta_0)^{i,j} j(\theta_0)^{k,l} \\
&\quad \times \{3T_{ijm}(\theta_0)j(\theta_0)^{m,n}T_{nkl}(\theta_0) + 2T_{ikm}(\theta_0)j(\theta_0)^{m,n}T_{nij}(\theta_0)\}, \\
\tilde{b} &= \frac{1}{12p(p+2)} j(\theta_0)^{i,j} j(\theta_0)^{k,l} \{3T_{ijkl}(\theta_0) - 6\alpha T_{i,jkl}(\theta_0) \\
&\quad + 3(1 + \alpha - 2\alpha^2)T_{ijm}(\theta_0)j(\theta_0)^{m,n}T_{nkl}(\theta_0) \\
&\quad + (5 + 6\alpha - 3\alpha^2)T_{ikm}(\theta_0)j(\theta_0)^{m,n}T_{nij}(\theta_0)\}, \\
\tilde{c} &= c.
\end{aligned}$$

Comparison of the coefficients in the expected and observed cases shows that these coefficients can be expressed in very similar ways in terms of tensors obtained from the expected and observed likelihood yokes, respectively. Many of the coefficients in the expected case coincide with the quantities formed by using the expected yoke instead of the observed likelihood yoke in the formulae for the coefficients in the observed case. The others differ only by terms which vanish for full exponential models. Thus, for full exponential models, $\mathcal{W}_e^\alpha = \tilde{\mathcal{W}}_e^{\alpha*}$ and $\tilde{\mathcal{W}}_e^\alpha = \mathcal{W}_e^{\alpha*}$.

In the case of a one-dimensional parametric statistical model, the coefficients a , b , c and \tilde{a} , \tilde{b} , \tilde{c} obtained from the expected likelihood yoke simplify to

$$a = \frac{(1 - 3\alpha)^2}{144} i(\theta_0)^{-3} T_{111}(\theta_0)^2,$$

$$b = \frac{1}{36} i(\theta_0)^{-2} \{3\tau_{1,1,1,1}(\theta_0) - 6(1 - 2\alpha)T_{1;111}(\theta_0) + (8 - 18\alpha - 9\alpha^2)i(\theta_0)^{-1}T_{111}(\theta_0)^2\},$$

$$c = \frac{1}{12} i(\theta_0)^{-2} \{-3\tau_{1,1,1,1}(\theta_0) - 12\tau_{11,1,1}(\theta_0) + 5i(\theta_0)^{-1}T_{111}(\theta_0)^2\}$$

and

$$\tilde{a} = \frac{(1 + 3\alpha)^2}{144} i(\theta_0)^{-3} T_{111}(\theta_0)^2,$$

$$\tilde{b} = \frac{1}{36} i(\theta_0)^{-2} \{3\tau_{1,1,1,1}(\theta_0) - 6(1 + \alpha)T_{1;111}(\theta_0) + (8 + 9\alpha - 9\alpha^2)i(\theta_0)^{-1}T_{111}(\theta_0)^2\},$$

$$\tilde{c} = c.$$

The coefficients a , b , c and \tilde{a} , \tilde{b} , \tilde{c} obtained from the observed likelihood yoke simplify to

$$a = \frac{(1 - 3\alpha)^2}{144} j(\theta_0)^{-3} T_{111}(\theta_0)^2,$$

$$b = \frac{1}{36} j(\theta_0)^{-2} \{3T_{1111;1}(\theta_0) + 12\alpha T_{1;111}(\theta_0) + (8 - 18\alpha - 9\alpha^2)j(\theta_0)^{-1}T_{111}(\theta_0)^2\},$$

$$c = \frac{1}{12} j(\theta_0)^{-2} \{3T_{1111;1}(\theta_0) + 12T_{11;11}(\theta_0) + 5j(\theta_0)^{-1}T_{111}(\theta_0)^2\}$$

and

$$\tilde{a} = \frac{(1 + 3\alpha)^2}{144} j(\theta_0)^{-3} T_{111}(\theta_0)^2,$$

$$\tilde{b} = \frac{1}{36} j(\theta_0)^{-2} \{3T_{1111;1}(\theta_0) - 6\alpha T_{1;111}(\theta_0) + (8 + 9\alpha - 9\alpha^2)j(\theta_0)^{-1}T_{111}(\theta_0)^2\},$$

$$\tilde{c} = c.$$

Acknowledgements

This paper is based on part of P.V. Larsen's doctoral thesis at the University of St Andrews. She is grateful to the University of St Andrews for a University Research Studentship. We are grateful also to Frank Critchley and Chris Jones for constructive discussions and for helpful comments on earlier drafts.

References

- Barndorff-Nielsen, O.E. and Cox, D.R. (1989) *Asymptotic Techniques for Use in Statistics*. London: Chapman & Hall.
- Barndorff-Nielsen, O.E. and Cox, D.R. (1994) *Inference and Asymptotics*. London: Chapman & Hall.
- Barndorff-Nielsen, O.E. and Hall, P. (1988) On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*, **75**, 374–378.
- Barndorff-Nielsen, O.E. and Jupp, P.E. (1997) Statistics, yokes and symplectic geometry. *Ann. Fac. Sci. Toulouse*, **6**, 389–427.
- Barndorff-Nielsen, O.E., Jupp, P.E. and Kendall, W.S. (1994) Stochastic calculus, statistical asymptotics, Taylor strings and phyla. *Ann. Fac. Sci. Toulouse*, **3**, 5–62.
- Blæsild, P. (1991) Yokes and tensors derived from yokes. *Ann. Inst. Statist. Math.*, **43**, 95–133.
- Breusch, R.S. and Schmidt, P. (1988) Alternative forms of the Wald test: How long is a piece of string? *Comm. Statist. Theory Methods*, **17**, 2789–2795.
- Cordeiro, G.M. and Ferrari, S.L.P. (1991) A modified score test statistic having chi-squared distribution to order n^{-1} . *Biometrika*, **78**, 573–582.
- Critchley, F., Marriott, P.K. and Salmon, M.H. (1993) Preferred point geometry and statistical manifolds. *Ann. Statist.*, **21**, 1197–1224.
- Critchley, F., Marriott, P.K. and Salmon, M.H. (1994) Preferred point geometry and the local differential geometry of the Kullback–Leibler divergence. *Ann. Statist.*, **22**, 1587–1602.
- Critchley, F., Marriott, P.K. and Salmon, M.H. (1996) On the differential geometry of the Wald test with nonlinear restrictions. *Econometrica*, **64**, 1213–1222.
- Gregory, A.W. and Veall, M.R. (1985) Formulating Wald tests of nonlinear restrictions. *Econometrica*, **53**, 1465–1468.
- Hayakawa, T. and Puri, M.L. (1985) Asymptotic expansions of the distribution of some test statistics. *Ann. Inst. Statist. Math.*, **37**, 95–108.
- Hougaard, P. (1982) Parametrizations of non-linear models. *J. Roy. Statist. Soc. Ser. B*, **44**, 244–252.
- Kass, R.E. (1984) Canonical parameterizations and zero parameter-effects curvature. *J. Roy. Statist. Soc. Ser. B*, **46**, 86–92.
- Le Cam, L. (1990) On the standard asymptotic confidence ellipsoids of Wald. *Int. Statist. Rev.*, **58**, 129–152.
- Lindsey, J.K. (1996) *Parametric Statistical Inference*. Oxford: Oxford University Press.
- Mora, M. (1992) Geometric expansions for the distributions of the score vector and the maximum likelihood estimator. *Ann. Inst. Statist. Math.*, **44**, 63–83.
- Phillips, P.C.B. and Park, J.Y. (1988) On the formulation of Wald tests of nonlinear restrictions. *Econometrica*, **56**, 1065–1083.
- Pierce, D.A. and Peters, D. (1992) Practical use of higher order asymptotics for multiparameter exponential families (with Discussion). *J. Roy. Statist. Soc. Ser. B*, **54**, 701–737.

Received August 2001 and revised June 2002