Information Geometry
https://doi.org/10.1007/s41884-023-00101-4

RESEARCH PAPER

A parameterisation-invariant modification of the score test

P. E. Jupp[1]

Received: 31 August 2022 / Revised: 10 December 2022 / Accepted: 4 February 2023

Abstract
The null distribution of the score test statistic is asymptotically chi-squared for large samples. The error in this approximation is improved greatly by a cubic modification. The coefficients of this cubic that are given in the literature depend on the parameterisation. This paper provides parameterisation-invariant versions of the coefficients, expresses them in terms of appropriate tensors, and provides geometric interpretations.

Keywords Bartlett correction · Generalised Bartlett correction · Interest parameter · Invariant Taylor expansion · Large-sample asymptotics · Likelihood yoke · Tensor

1 Introduction

A common activity in statistics is that of testing the null hypothesis, H_0, that the true value of the parameter ω lies in a specified subspace of the parameter space Ω. The two main general tests are the likelihood ratio test (LRT) and the score test. The LRT rejects H_0 for large values of $w = 2\left\{l(\hat{\omega}; x_1, \ldots, x_n) - l(\tilde{\omega}; x_1, \ldots, x_n)\right\}$, where $l(\cdot; x_1, \ldots, x_n)$ denotes the log-likelihood based on observations $x_1, \ldots, x_n$, and $\hat{\omega}$ and $\tilde{\omega}$ are the maximum likelihood estimate and the restricted maximum likelihood estimate under H_0, respectively. The score test rejects H_0 for large values of

$$S = \tilde{U}_h^\top \tilde{i}_h^{-1} \tilde{U}_h, \tag{1}$$

where U_h is the score for the interest parameter, ${i_h}^{-1}$ is the interest part of the inverse Fisher information, $U_h^\top$ denotes the transpose of U_h, and each tilde indicates evaluation at $\tilde{\omega}$.

Communicated by Hiroshi Matsuzoe.

✉ P. E. Jupp
pej@st-andrews.ac.uk

[1] School of Mathematics and Statistics, University of St Andrews, St Andrews KY16 9SS, UK

Under mild regularity conditions and under independent sampling, the large-sample asymptotic null distributions of w and S are χ^2_p with error of order $O(n^{-1})$, where p is the dimension of the interest parameter. For w, there is a Bartlett adjusted version, w^*, of w given by

$$w^* = w(1 + R/p) \tag{2}$$

for some constant R and so that the null distribution of w^* is χ^2_p with error of order $O(n^{-2})$ [3, 4]. The scalar R can be expressed in terms of some tensors [3, 6] that arise from the geometry. For S there is no analogous linear Bartlett adjustment but there is a cubic modification [9] of S such that its null distribution is χ^2_p with error of order $O(n^{-3/2})$. The coefficients of the cubic are linear functions of coefficients in the expansion [11] to order $O(n^{-1})$ of the moment generating function of S. These coefficients (and so the cubic modification) depend on the choice of parameterisation of the nuisance parameters, i.e., on the way in which the parameter space is written locally as a product of the spaces of interest and nuisance parameters. Even after correction of a misprint noted by [9], the coefficients of the cubic given in [11] are not invariant under re-parameterisation [10]. Further, there are no obvious geometric interpretations of the coefficients. For the case of simple null hypotheses, there is [8] a parameterisation-invariant version, $S^\ddagger$, of S such that the null distribution of $S^\ddagger$ is χ^2_p with error of order $O(n^{-2})$. Whereas the cubic correction, S^*, of S introduced in Sect. 3.3 below is a cubic function of S, $S^\ddagger = (\tilde{U}_h^\ddagger)^\top \tilde{i}_h^{-1} \tilde{U}_h^\ddagger$, where $\tilde{U}_h^\ddagger$ is a cubic function of $\tilde{U}_h$. Even in some simple models (such as that in [14, Sect. 3]), the cubic giving $\tilde{U}_h^\ddagger$ in terms of $\tilde{U}_h$ can be quite complicated. There are no obvious geometric interpretations of the coefficients of this cubic.

The aim of this paper is to provide a parameterisation-invariant expansion to order $O(n^{-1})$ of S in which the coefficients have geometric interpretations. A cubic correction, S^*, of S is introduced, such that the null distribution of S^* is χ^2_p with error of order $O(n^{-2})$. Because two serious disadvantages of index notation are (i) it is vulnerable to misprints, (ii) it can obscure concepts by concentrating on the details of calculations, the approach here largely avoids explicit parameterisations and the use of index notation. For readers who prefer index notation, Appendix A contains expressions in that language for the coefficients of the cubic.

Section 2 recalls material on yokes, introduces fibred yokes, and shows how they give rise to decomposition of tensors. In Sect. 3 the asymptotic moment generating function of S is derived, the coefficients of the cubic giving S^* are given, and these coefficients are related to appropriate tensors.

2 Yokes and fibred yokes

An appropriate geometric setting for parametric models in which nuisance parameters can be present is that of submersions from one smooth manifold to another. More precisely, $\pi : \Omega \to \Psi$ is a smooth map from the full parameter space, Ω, to the space, Ψ, of parameters of interest, and at each point ω of Ω the tangent map π_* maps the tangent space $T\Omega_\omega$ *onto* $T\Psi_{\pi(\omega)}$. The submersion condition implies that each fibre $\pi^{-1}(\psi)$ is a submanifold of Ω and that around each ω small portions of Ω look like $\Psi \times \pi^{-1}(\pi(\omega))$ with π being identified locally with the projection of $\Psi \times \pi^{-1}(\pi(\omega))$ onto Ψ. Nevertheless, in general Ω is not such a product and it is conceptually not helpful to think of Ω in this way.

2.1 Yokes

The coordinate-free definition of a yoke is as follows. For a vector field X on a manifold Ω, define the vector fields $\bar{X}$ and $\bar{X}'$ on $\Omega \times \Omega$ by $\bar{X} = (X, 0)$ and $\bar{X}' = (0, X)$, i.e.,

$$\begin{aligned} Tp_1(\bar{X}) &= X, \qquad & Tp_2(\bar{X}) &= 0, \\ Tp_1(\bar{X}') &= 0, \qquad & Tp_2(\bar{X}') &= X, \end{aligned}$$

where $p_k : \Omega \times \Omega \to \Omega$ is the projection onto the k^{th} factor for $k = 1, 2$. Then, for vector fields X and Y on Ω and a smooth function $g : \Omega \times \Omega \to \mathbb{R}$, we define $g(X|Y) : \Omega \to \mathbb{R}$ by

$$g(X|Y)(\omega) = \bar{X}\bar{Y}' g(\omega, \omega).$$

A yoke on Ω may now be characterised as a smooth function $g : \Omega \times \Omega \to \mathbb{R}$ such that

(i) $\bar{X} g(\omega, \omega) = 0$ for all ω in Ω,
(ii) the (0,2)-tensor $(X, Y) \mapsto g(X|Y)$ is non-singular.

An alternative way of expressing (i) and (ii) is that on the diagonal $\Delta_\Omega = \{(\omega, \omega) : \omega \in \Omega\}$,

(i) $d_1 g = 0$,
(ii) $d_1 d_2 g$ is non-singular,

where d_1 and d_2 denote exterior differentiation along the first and second factor, respectively, in $\Omega \times \Omega$.

The two main yokes of interest in statistics are the likelihood yokes. Consider a parametric statistical model with parameter space Ω, sample space $\mathcal{X}$ and log-likelihood function $l : \Omega \times \mathcal{X} \to \mathbb{R}$. The *expected likelihood yoke* on Ω is the function g on $\Omega \times \Omega$ given by

$$g(\omega, \omega') = E_{\omega'}[l(\omega; x) - l(\omega'; x)]. \tag{3}$$

Suppose that an auxiliary statistic a is given, such that the statistic $(\hat{\omega}, a)$ is minimal sufficient for ω, where $\hat{\omega}$ denotes the maximum likelihood estimator. Then the corresponding *observed likelihood yoke* on Ω is the function g on $\Omega \times \Omega$ given by

$$g(\omega, \omega') = l(\omega; \omega', a) - l(\omega'; \omega', a). \tag{4}$$

Properties and applications of expected and observed likelihood yokes can be found in [1, 3].

A key property of yokes is that they give rise naturally to preferred coordinate charts (called *extended normal coordinates*) taking values in appropriate cotangent spaces. Given any point ω of Ω, the function Γ_ω from Ω to the cotangent space $T^*\Omega_\omega$ to Ω at ω is defined by

$$\Gamma_\omega(\omega') = d_1 g(\omega, \omega'). \tag{5}$$

In terms of local coordinates $\omega^1, \ldots, \omega^d$ on Ω,

$$\Gamma_\omega(\omega') = \frac{\partial g(\omega, \omega')}{\partial \omega^u} d\omega^u,$$

where the Einstein summation convention is used. It follows from property (i) of a yoke that $\Gamma_\omega(\omega) = 0$ and from property (ii) that the restriction $\Gamma_\omega|U$ of Γ_ω to some neighbourhood U of ω in Ω is a coordinate chart on U taking values in $T^*\Omega_\omega$. Note that the space $T^*\Omega_\omega$ depends on ω. It has been customary [3, Sect. 5.6], [6, Sect. 4], [16] to use the metric given by the yoke to 'raise' the Γ_ω, in order to obtain extended normal coordinates with values in the tangent space $T\Omega_\omega$ rather than in its dual, the cotangent space $T^*\Omega_\omega$. The Γ_ω defined in (5) are used here because they can be regarded as more basic. In the language of strings, the coordinate expressions for the 'raised' versions of the derivatives of Γ_ω form the costring $\overset{-1}{g\!\!/}$ [6].

For any smooth function f on Ω, the composition $f \circ \Gamma_\omega^{-1}$ is a function on an open neighbourhood of 0 in the vector space $T^*\Omega_\omega$, and so its derivatives are symmetric tensors on $T^*\Omega_\omega$. Combining these tensors with Γ_ω gives an *invariant Taylor expansion* (a parameterisation-invariant analogue of a Taylor expansion) of f. Expressions in index notation for ('lowered' versions of) these invariant Taylor expansions are given in [5, Sect. 3.3], [3, Sect. 5.6], [16, Sect. 4]. Similarly, for any smooth function h on $\Omega \times \Omega$, the composition $h \circ (\Gamma_\omega^{-1} \times \Gamma_\omega^{-1})$ is a function on an open neighbourhood of 0 in the vector space $T^*\Omega_\omega \times T^*\Omega_\omega$, and so its derivatives are symmetric tensors on $T^*\Omega_\omega \otimes T^*\Omega_\omega$. In the language of strings, these tensors are said to be obtained by intertwining [1]. In particular, Taylor expansion of g in the corresponding product coordinate charts on a neighbourhood of (ω, ω) in $\Omega \times \Omega$ yields a family of tensors $T_{r_1,\ldots,r_p;s_1,\ldots,s_q}$ on ω [6].

Remark 1 Extended normal coordinates, Γ_ω, can be defined also in the more general setting of *pre-contrast functions*, meaning functions $h : \Omega \times \Omega \to T^*\Omega$ such that

(o) $h(\omega, \omega') \in T^*\Omega_\omega$,
(i) $h(\omega, \omega) = 0$,
(ii) $d_2 h$ is non-degenerate on the diagonal, Δ_Ω, where d_2 denotes the exterior derivative along $\{\omega\} \times \Omega$.

(In the language of vector bundles, h is a section of the pull-back of the cotangent bundle of Ω by the projection $\pi_1 : \Omega \times \Omega \to \Omega$ onto the first factor, such that $h = 0$ on the diagonal and its derivative is non-degenerate there.) The original definition

[12] of pre-contrast functions required the restriction of $-d_2 h$ to the diagonal to be a semi-Riemannian metric on Ω.

The general mathematical concept that underlies the results in this paper is that of a *fibred yoke*, i.e., a submersion $\pi : \Omega \to \Psi$, together with a yoke on Ω. In the current context, π maps parameters to interest parameters, and the yoke is a likelihood yoke (3) or (4).

2.2 Decomposition of tangent spaces

In the tangent space $T\Omega_\omega$ to Ω at ω the *vertical subspace* V_ω is defined as $V_\omega = \{X \in T\Omega_\omega : \pi_*(X) = 0\}$. Given a Riemannian metric ϕ on Ω, the *horizontal subspace* H_ω is the orthogonal complement of V_ω in $T\Omega_\omega$. Thus ϕ decomposes $T\Omega_\omega$ as the orthogonal direct sum

$$T\Omega_\omega = V_\omega \oplus H_\omega. \tag{6}$$

The decomposition (6) varies smoothly with ω, in the sense that $\omega \mapsto (V_\omega, H_\omega)$ is a smooth map from Ω to $V_q(T\Omega) \times V_p(T\Omega)$, where $V_r(T\Omega)$ denotes the manifold $\{(\omega, E_\omega) : E_\omega \text{ is an } r\text{-dimensional subspace of } T\Omega_\omega\}$, and p and q are the dimensions of the interest and nuisance parameters, respectively. The smoothness of the decomposition (6) implies that Y_h, Y_v, Y_{hv}, and Y_{vv} defined in Subsection 3.1 depend smoothly on ω, and so, under mild regularity conditions, the tensors defined in (9) below exist. The tangent mapping π_* identifies H_ω with $T\Psi_{\pi(\omega)}$.

The inner product $\pi_\omega\phi$ on $T\Psi_{\pi(\omega)}$ is defined by

$$\pi_\omega\phi(X, Y) = \phi(\tilde{X}, \tilde{Y}) \quad X, Y \in T\Psi_{\pi(\omega)},$$

where $\tilde{X}$ and $\tilde{Y}$ are the horizontal lifts to $T\Omega_\omega$ of X and Y, i.e., they are the unique elements of H_ω such that $\pi_*(\tilde{X}) = X$ and $\pi_*(\tilde{Y}) = Y$. The dual of the decomposition (6) of the tangent space $T\Omega_\omega$ to Ω at ω is the decomposition

$$T^*\Omega_\omega = V_\omega^* \oplus H_\omega^* \tag{7}$$

of the cotangent space $T^*\Omega_\omega$ to Ω at ω. Taking the r-fold tensor product of the decomposition (7) of $T^*\Omega_\omega$ leads to the decomposition

$$\otimes^r T^*\Omega_\omega = \oplus_{s=0}^r \left((\otimes^s V_\omega^*) \otimes \left(\otimes^{r-s} H_\omega^*\right)\right) \tag{8}$$

of the space of r-fold tensors on $T^*\Omega_\omega$.

The projection of the score onto H_ω^* using the decomposition (7) is the *horizontal score*, U_h, used in (1). It is the score for the interest parameter, ψ, and is also known as the *orthogonal score* [13, 17].

3 Higher-order behaviour of *S*

3.1 Tensors from log-likelihood derivatives

Denote by Z_1, Z_2, Z_3 the 1st, 2nd and 3rd derivatives of the log-likelihood, centred and scaled by $n^{-1/2}$ to have order $O_p(1)$. Expressing Z_1, Z_2, Z_3 in the functions Γ_ω around ω given by (5) with the expected likelihood yoke (3) yields random tensors Y_1, Y_2, Y_3. Decomposing Y_1, Y_2, Y_3 by (8) gives Y_h in H^*_ω, Y_v in V^*_ω, Y_{hv} in $H^*_\omega \otimes V^*_\omega$, Y_{vv} in $\otimes^2 V^*_\omega$ and Y_{hvv} in $H^*_\omega \otimes (\otimes^2 V^*_\omega)$. The tensors $\tau_{h,h,h}$ in $\otimes^3 H^*_\omega$, $\tau_{h,h,v}$ in $(\otimes^2 H^*_\omega) \otimes V^*_\omega$, $\tau_{h,v,v}$ in $H^*_\omega \otimes (\otimes^2 V^*_\omega)$, $\tau_{hv,hv}$ in $\otimes^2(H^*_\omega \otimes V^*_\omega)$, $\tau_{h,h,vv}$ in $(\otimes^2 H^*_\omega) \otimes (\otimes^2 V^*_\omega)$, $\tau_{h,v,hv}$ in $\otimes^2(H^*_\omega \otimes V^*_\omega)$ and $\tau_{h,h,h,h}$ in $\otimes^4 H^*_\omega$ are defined by

$$
\begin{aligned}
\tau_{h,h,h} = E[\otimes^3 Y_h], \quad \tau_{h,h,v} = E[(\otimes^2 Y_h) \otimes Y_v], \quad \tau_{h,v,v} = E[Y_h \otimes (\otimes^2 Y_v)],\\
\tau_{hv,hv} = E[Y_{hv} \otimes Y_{hv}], \ \tau_{h,h,vv} = E[Y_h \otimes Y_h \otimes Y_{vv}], \ \tau_{h,v,hv} = E[Y_h \otimes Y_v \otimes Y_{hv}],\\
\tau_{h,h,h,h} = E[\otimes^4 Y_h]. \qquad (9)
\end{aligned}
$$

Remark 2 The tensors (9) can be obtained from the expected yoke (3). There are analogous tensors [3, Sect. 5.5] arising from the observed likelihood yoke (4). Under ordinary repeated sampling, corresponding tensors differ by $O(n^{-1/2})$.

3.2 Moment generating function of *S*

One way [3, Sect. 5.3] of deriving the constant R in the expression (2) for w^* is based on expanding w to order $O(n^{-1})$ as a quartic in the score. There is an analogous expansion of S as

$$S = S_0 + n^{-1/2} S_1 + n^{-1} S_2 + O(n^{-3/2}),$$

where S_0, S_1, S_2 are $O_p(n^{-1})$, S_0 is a homogeneous quadratic in Y_1, S_1 is a homogeneous cubic in Y_1, Y_2, and S_2 is a homogeneous quartic in Y_1, Y_2, Y_3. Calculation of some low-order moments of products of S_0, S_1 and S_2 leads to the following theorem.

Theorem 1 *Suppose that (a) the sample space is continuous, (b) the log-likelihood function is finite and its derivatives of order 4 or less are continuous in some neighbourhood of ω, (c) the Fisher information at ω is non-singular. Then the moment generating function $M_S(t)$ of S has the form*

$$M_S(t) = (1-2t)^{-\frac{p}{2}} \left\{ 1 + \frac{1}{24n} \left(A_1 d + A_2 d^2 + A_3 d^3 + O(d^4) \right) \right\} + O(n^{-3/2}) \qquad (10)$$

where $d = 2t/(1-2t)$ and

$$A_1 = 12\,\mathrm{tr}_h\,\mathrm{tr}_v\,(\tau_{hv,hv}) + 3\,\langle \mathrm{tr}_v\,(\tau_{h,v,v}), \mathrm{tr}_h\,(\tau_{h,h,h})\rangle_h$$
$$+ 6\,\|\tau_{h,v,v}\|^2 + 6\,\mathrm{tr}_h\,\mathrm{tr}_v(\tau_{h,h,vv}) + 36\,\mathrm{tr}_h\,\mathrm{tr}_v(\tau_{h,v,hv}) + 6\,\|\mathrm{tr}_v\,\tau_{h,v,v}\|^2, \quad (11)$$
$$A_2 = 3\,\mathrm{tr}_h\,\mathrm{tr}_h\,(\tau_{h,h,h,h}) - 6\|\tau_{h,h,v}\|^2 - 3\|\mathrm{tr}_h\,(\tau_{h,h,v})\|_v^2$$
$$- 6\,\langle \mathrm{tr}_v\,(\tau_{h,v,v}), \mathrm{tr}_h\,(\tau_{h,h,v})\rangle_h\,, \quad (12)$$
$$A_3 = 3\,\|\mathrm{tr}\,(\tau_{h,h,h})\|^2 + 2\,\|\tau_{h,h,h}\|^2, \quad (13)$$

where tr_h *and* tr_v *indicate traces taken over pairs of factors in* H_ω^* *and* V_ω^**, respectively, while inner products and norms on the tensor spaces* $\otimes H_\omega^*$*, etc. are those given by tensor products of inverse Fisher information.*

If the null hypothesis, H_0, is simple then

$$A_1 = 0, \quad (14)$$
$$A_2 = 3\,\mathrm{tr}\,\mathrm{tr}\,(\tau_4)\,, \quad (15)$$
$$A_3 = 3\,\|\mathrm{tr}\,(\tau_3)\|^2 + 2\,\|\tau_3\|^2, \quad (16)$$

where $\tau_3 = \tau_{h,h,h}$, $\tau_4 = \tau_{h,h,h,h}$, and the expressions given in [11, (3)] agree with (14)–(16) Further, in this case of a simple H_0, the constant R in the definition (2) of the Bartlett adjusted version w^* of w can be expressed as

$$R = \frac{1}{12}\left\{12\,\mathrm{tr}\,\mathrm{tr}\,(\tau_{2,2}) + A_2 + A_3\right\}$$

with A_2 and A_3 as in (15)–(16) and $\tau_{2,2}$ in $\otimes^4 T^*\Omega_\omega$ defined with components $\mathcal{T}_{ij;kl}$ in [6, (5.22)]. There is also an expression [3, 6] for R in terms of analogous tensors (mentioned in Remark 2) arising from the observed likelihood yoke (4).

3.3 Cubic modification of *S*

Put

$$c = \frac{A_1 - A_2 + A_3}{12p}, \qquad b = \frac{A_2 - 2A_3}{12p(p+2)}, \qquad a = \frac{A_3}{12p(p+2)(p+4)},$$

where p is the dimension of Ψ, and define the cubic modification S^* of S by

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

Then [9] the null distribution of S^* is χ_p^2 with error of order $O(n^{-3/2})$. A slight extension of the symmetry argument in [4] for the Bartlett-corrected likelihood ratio test shows that the error is of order $O(n^{-2})$.

Acknowledgements I am indebted to Eleanor Emberson for discussions and for detailed calculations in [10] that led to many of the results presented here. I am grateful to Pia Veldt Larsen for discussions on the null asymptotic distributions of the score statistic and geometric Wald statistics. I thank two referees for their helpful comments.

Author Contributions Not applicable.

Funding This research was not funded.

Data availabilty Not applicable.

Declarations

Conflict of interest There are no competing interests.

Consent to participate Not applicable.

Consent for publication Not applicable.

Appendix A Coefficients A_1–A_3 in terms of cumulants of log-likelihood derivatives

The coefficients A_1–A_3 can be expressed in terms of the cumulants of log-likelihood derivatives (for a single observation). In terms of local coordinates $\omega^1, \ldots, \omega^{p+q}$ on Ω, these cumulants have components

$$
\begin{aligned}
\kappa_{ij} &= E\left[\frac{\partial^2 l}{\partial\omega^i \partial\omega^j}(\omega; x)\right], \\
\kappa_{ijk} &= E\left[\frac{\partial^3 l}{\partial\omega^i \partial\omega^j \partial\omega^k}(\omega; x)\right], \\
\kappa_{i,j} &= E\left[\frac{\partial l}{\partial\omega^i}(\omega; x)\frac{\partial l}{\partial\omega^j}(\omega; x)\right], \\
\kappa_{i,j,k} &= E\left[\frac{\partial l}{\partial\omega^i}(\omega; x)\frac{\partial l}{\partial\omega^j}(\omega; x)\frac{\partial l}{\partial\omega^k}(\omega; x)\right],
\end{aligned}
$$

etc.

Suppose that $\omega^1, \ldots, \omega^{p+q}$ are chosen such that $\omega^1, \ldots, \omega^p$ are interest parameters, whereas $\omega^{p+1}, \ldots, \omega^{p+q}$ are nuisance parameters. Let

$$K = \begin{pmatrix} K_{1,1} & K_{1,2} \\ K_{2,1} & K_{2,2} \end{pmatrix}$$

be the $(p+q) \times (p+q)$ matrix of the $\kappa_{i,j}$, partitioned into blocks corresponding to the interest and nuisance parameters, respectively. Put

$$\begin{aligned} A &= \begin{pmatrix} 0 & 0 \\ 0 & K_{2,2}^{-1} \end{pmatrix}, \\ M &= K^{-1} - A. \end{aligned} \tag{A1}$$

Then expressions (11–13) can be written in index notation as

$$\begin{aligned} A_1 &= 12\,(\kappa_{ij,kl} - \kappa^{m,n}\kappa_{m,ij}\kappa_{n,kl})m^{ik}m^{jl} \\ &\quad + 3\,\kappa_{i,j,k}\kappa_{l,m,n}m^{il}a^{jk}a^{mn} + 6\,\kappa_{i,j,k}\kappa_{l,m,n}m^{il}a^{jm}a^{kn} \\ &\quad + 6\,(\kappa_{i,j,kl} - \kappa^{m,n}\kappa_{m,i,j}\kappa_{n,kl})m^{ij}a^{kl} \\ &\quad + 36\,(\kappa_{i,j,kl} - \kappa^{m,n}\kappa_{m,i,j}\kappa_{n,kl})m^{ik}a^{jl} \\ &\quad + 6\,\kappa_{i,j,k}\kappa_{l,m,n}a^{il}m^{jk}a^{mn} \\ A_2 &= 3\,\kappa_{i,j,k,l}m^{ij}m^{kl} - 6\,\kappa_{i,j,k}\kappa_{l,m,n}a^{il}m^{jm}m^{kn} \\ &\quad - 3\,\kappa_{i,j,k}\kappa_{l,m,n}a^{il}m^{jk}m^{mn} - 6\,\kappa_{i,j,k}\kappa_{l,m,n}m^{il}a^{jk}m^{mn} \\ A_3 &= 3\,\kappa_{i,j,k}\kappa_{l,m,n}m^{il}m^{jk}m^{mn} + 2\,\kappa_{i,j,k}\kappa_{l,m,n}m^{il}m^{jm}m^{kn}, \end{aligned}$$

where indices run over $1, \ldots, p+q$, the $\kappa^{i,j}$ are the elements of K^{-1}, and the Einstein summation convention is used.

Appendix B Proof of Theorem 1

The proof of Theorem 1 proceeds along the lines of the derivation of the expression for the Bartlett correction factor given in [15, Sect. 7.4] and [3, Sect. 5.3]. Only an outline of the proof is given here; full details can be found in [10].

Step 1: S in terms of polynomials in Y_v, Y_h, Y_{hv}, Y_{vv} and Y_{hvv}

Ordinary Taylor series expansion (in any coordinate system on the full parameter space Ω) of Z_1 and i_h^{-1} about ω gives (in index notation)

$$\begin{aligned} \tilde{Z}_i &= Z_i + \kappa_{ij}\tilde{\delta}^j + n^{-1/2}\left(Z_{ij}\tilde{\delta}^j + \frac{1}{2}\kappa_{ijk}\tilde{\delta}^j\tilde{\delta}^k\right) \\ &\quad + n^{-1}\left(\frac{1}{2}Z_{ijk}\tilde{\delta}^j\tilde{\delta}^k + \frac{1}{6}\kappa_{ijkl}\tilde{\delta}^j\tilde{\delta}^k\tilde{\delta}^l\right) + O(n^{-3/2}) \end{aligned} \tag{B1}$$

and

$$\tilde{\kappa}^{i,j} = \kappa^{i,j} - n^{-1/2}\kappa^{i,r}\left[\left(\frac{\partial}{\partial\omega^a}\kappa_{r,s}\right)\tilde{\delta}^a + n^{-1/2}\frac{1}{2}\left(\frac{\partial^2}{\partial\omega^a\partial\omega^b}\kappa_{r,s}\right)\tilde{\delta}^a\tilde{\delta}^b\right]\kappa^{s,j}$$
$$+ n^{-1}\kappa^{i,r}\left(\frac{\partial}{\partial\omega^a}\kappa_{r,s}\right)\tilde{\delta}^a\kappa^{s,t}\left(\frac{\partial}{\partial\omega^b}\kappa_{t,u}\right)\tilde{\delta}^b\kappa^{u,j} + O(n^{-3/2}), \tag{B2}$$

where $\tilde{\delta}^i = n^{1/2}(\tilde{\omega} - \omega)^i$. Since $\tilde{Z}_i = 0$ if ω_i is a nuisance parameter, (B1) can be solved to give Z_i (up to $O(n^{-3/2})$) as a cubic in the $\tilde{\delta}^j$. Substituting (B1) and (B2) in (1) then gives S (up to $O(n^{-3/2})$) as a cubic in the Z_j. For general coordinate systems the coefficients of this cubic are very complicated expressions in the first four cumulants of the score but if the coordinate charts Γ_ω are used then the coefficients take a much simpler form and

$$S = S_0 + n^{-1/2}S_1 + n^{-1}S_2 + O(n^{-3/2}),$$

where S_0, S_1, S_2 are polynomials (of degrees 2, 3 and 4, respectively) in Y_v, Y_h, Y_{hv}, Y_{vv} and Y_{hvv}.

Step 2: The moment generating function of S.

The randomness in S comes from Y, where $Y = (Y_1, Y_2, Y_3)$. An approximation to order $O(n^{-1})$ to the probability density function of Y is obtained by Edgeworth expansion in terms of tensorial Hermite polynomials [2, Sect. 5.7] of orders 3 and 4. The regularity conditions in Theorem 1 ensure that this Edgeworth expansion is valid (see [7, Sect. 5]). Then the moment generating function M_S of S satisfies

$$M_S(t) = \frac{|2\pi V|^{-1/2}}{|2\pi W|^{-1/2}}\int |2\pi W|^{-1/2}\exp\left\{-\frac{1}{2}y^\top W^{-1}y\right\}P(y)dy + O(n^{-3/2}), \tag{B3}$$

where V is the variance matrix of Y, $W = (I - 2tVU)^{-1}V$ with

$$U = \begin{pmatrix} M & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

M being given by (A1), and P is a function of S_0, S_1, S_2 and tensorial Hermite polynomials in Y_1, Y_2, Y_3. Equation (B3) can be written in terms of moments of S_0, S_1, S_2 and the tensorial Hermite polynomials. Calculation of these moments, together with some manipulation, then yields (10) and (11)–(13).

References

1. Barndorff-Nielsen, O.E., Blæsild, P.: Strings: mathematical theory and statistical examples. Proc. R. Soc. Lond. A **411**, 155–176 (1987)

2. Barndorff-Nielsen, O.E., Cox, D.R.: Asymptotic techniques for use in statistics. Chapman & Hall, London (1989)
3. Barndorff-Nielsen, O.E., Cox, D.R.: Inference and asymptotics. Chapman & Hall, London (1994)
4. Barndorff-Nielsen, O.E., Hall, P.: On the level-error after Bartlett adjustment of the likelihood ratio statistic. Biometrika **75**, 374–388 (1988)
5. Barndorff-Nielsen, O.E., Jupp, P.E., Kendall, W.S.: Stochastic calculus, statistical asymptotics, Taylor strings and phyla. Ann. Fac. Sci. Toulouse **3**, 5–62 (1994)
6. Blæsild, P.: Yokes and tensors derived from yokes. Ann. Inst. Stat. Math. **43**, 95–113 (1991)
7. Chandra, T.K., Ghosh, J.K.: Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables. Sankhyā **41**, 22–47 (1979)
8. Chandra, T.K., Mukerjee, R.: Bartlett-type modification for Rao's efficient score statistic. J. Multivariate Anal. **36**, 103–112 (1991)
9. Cordeiro, G.M., Ferrari, S.L.P.: A modified score test statistic having chi-squared distribution to order n^{-1}. Biometrika **78**, 573–582 (1991)
10. Emberson, E.A.: The asymptotic distribution and robustness of the likelihood ratio and score test statistics. Doctoral dissertation, University of St Andrews (1994)
11. Harris, P.: An asymptotic expansion for the null distribution of the efficient score statistic. Biometrika **72**, 653–659 (1985) Correction: Biometrika **74**, 667 (1987)
12. Henmi, M., Matsuzoe, H.: Geometry of pre-contrast functions and non-conservative estimating functions. In: International Workshop on Complex Structures, Integrability and Vector Fields. AIP Conf. Proc. 1340, pp. 32–41. https://doi.org/10.1063/1.3567122 (2011)
13. Hudson, S., Vos, P.W.: Marginal information for expectation parameters. Canad. J. Statist. **28**, 875–886 (2000)
14. Jupp, P.E.: Modifications of the Rayleigh and Bingham tests for uniformity of directions. J. Multivariate Anal. **77**, 1–20 (2001)
15. McCullagh, P.: Tensor methods in statistics. Chapman & Hall, London (1987)
16. Pace, L., Salvan, A.: The geometric structure of the expected/observed likelihood expansions. Ann. Inst. Stat. Math. **46**, 649–666 (1994)
17. Zhu, Y., Reid, N.: Information, ancillarity and sufficiency in the presence of nuisance parameters. Canad. J. Statist. **22**, 111–123 (1994)