Ann. Inst. Statist. Math.
Vol. 45, No. 4, 759–771 (1993)

AN EXTENSION OF THE CONDITIONAL LIKELIHOOD RATIO TEST TO THE GENERAL MULTIPARAMETER CASE

RAHUL MUKERJEE

Indian Institute of Management, Joka, Diamond Harbour Road, Post Box No. 16757, Alipore Post Office, Calcutta 700 027, India

(Received February 12, 1992; revised December 22, 1992)

Abstract. In a set-up, where both the interest parameter and the nuisance parameter are possibly multi-dimensional and global parametric orthogonality may not hold, we suggest a test that is superior to the usual likelihood ratio test with regard to second-order local maximinity. The test can be motivated from the principles of conditional and adjusted likelihood.

Key words and phrases: Adjusted likelihood, conditional likelihood, contiguous alternative, local maximinity, parametric orthogonality, power, second order.

1. Introduction

Cox and Reid (1987) pioneered the idea of conditional likelihood as an effective means for handling nuisance parameters, discussed many interesting features of the same and raised several open issues; see also Cox (1988) for a very informative further discussion. One of these open issues related to an extension of their ideas to a general multiparameter set-up where the interest parameter and the nuisance parameter are both possibly multi-dimensional and global parametric orthogonality may not hold. The present work attempts to settle this problem to some extent.

With one-dimensional interest parameter and under global parametric orthogonality, it is known (Mukerjee (1992*a*)) that the conditional likelihood ratio test of Cox and Reid (1987) is superior to the usual likelihood ratio test in terms of (a) second-order local maximinity, and (b) proximity to the second-order power function attainable by a likelihood ratio test with known nuisance parameter. In a general multiparameter set-up, here we suggest a test which is shown to be superior to the likelihood ratio test in the sense (a) even without parametric orthogonality and in the sense (b) under parametric orthogonality. It is seen that the test proposed here can be motivated from the principle of conditional likelihood and also from that of adjusted likelihood (McCullagh and Tibshirani (1990)) which represents another important technique for handling nuisance parameters.

For further significant results in this general area of research, we refer to Barndorff-Nielsen (1986), Liang (1987), Conniffe (1990), Godambe (1991) and the references therein.

2. Notation and preliminaries

Let $\{X_i\}$, $i \geq 1$, be a sequence of independent and identically distributed random variables with common density $f(x;\theta)$, where $\theta = (\theta_{(1)}, \ldots, \theta_{(r)})'$ belongs to $\mathcal{R}^r$ or some open subset thereof and $r \geq 2$. Let $\theta_1 = (\theta_{(1)}, \ldots, \theta_{(p)})'$ be the parameter of interest and $\theta_2 = (\theta_{(p+1)}, \ldots, \theta_{(r)})'$ be the nuisance parameter, where $1 \leq p < r$. Both θ_1 and θ_2 are possibly multi-dimensional. Consider the null hypothesis $H_0 : \theta_1 = \theta_{10}$, where $\theta_{10} = (\theta_{(10)}, \ldots, \theta_{(p0)})'$ is a known $p \times 1$ vector, against the alternative $\theta_1 \neq \theta_{10}$. For power studies, we shall consider contiguous alternatives of the form $\theta_{1n} = \theta_{10} + n^{-1/2}d$, where $d = (d_1, \ldots, d_p)'$ and n is the sample size. All formal expansions used in this paper are over a set with P_{θ_{1n},θ_2}-probability $1 + o(n^{-1/2})$; see Chandra and Ghosh (1979, 1980). We also assume standard regularity conditions.

For $1 \leq i, j, u \leq r$, let D_i denote the operator of partial differentiation with respect to $\theta_{(i)}$ and define

$$
\begin{aligned}
I_{(ij)}(\theta) &= E_\theta[\{D_i \log f(X;\theta)\}\{D_j \log f(X;\theta)\}],\\
K_{iju}(\theta) &= E_\theta\{D_i D_j D_u \log f(X;\theta)\},\\
K_{i,ju}(\theta) &= E_\theta[\{D_i \log f(X;\theta)\}\{D_j D_u \log f(X;\theta)\}],\\
K_{i,j,u}(\theta) &= E_\theta[\{D_i \log f(X;\theta)\}\{D_j \log f(X;\theta)\}\{D_u \log f(X;\theta)\}],\\
H_{(i)}(\theta) &= n^{-1/2} \sum_{s=1}^{n} D_i \log f(X_s;\theta),\\
H_{(ij)}(\theta) &= n^{-1/2} \sum_{s=1}^{n} \{D_i D_j \log f(X_s;\theta) + I_{(ij)}(\theta)\}.
\end{aligned}
$$

Among the expectations defined above, only those which have been useful in the subsequent derivation are assumed to exist. It is also assumed that these are smooth functions of θ. Let $I(\theta) = ((I_{(ij)}(\theta)))$ be the $r \times r$ per observation Fisher's information matrix at θ which is assumed to be positive definite at each θ. Define the $r \times 1$ vector $H(\theta) = (H_{(1)}(\theta), \ldots, H_{(r)}(\theta))'$ and the $r \times r$ matrix $Q(\theta) = ((H_{(ij)}(\theta)))$. Also, let $L(\theta)$ be an $r \times r^2$ matrix with its i-th row given by $(K_{i11}(\theta), \ldots, K_{i1r}(\theta), \ldots, K_{ir1}(\theta), \ldots, K_{irr}(\theta))$, $1 \leq i \leq r$.

For the subsequent development, we shall require the partitioned forms of the vectors and matrices defined above. To that effect, let $H(\theta) = (H_1(\theta), H_2(\theta))'$, and

$$
Q(\theta) = \begin{pmatrix} Q_{11}(\theta) & Q_{12}(\theta) \\ Q_{21}(\theta) & Q_{22}(\theta) \end{pmatrix}, \qquad I(\theta) = \begin{pmatrix} I_{11}(\theta) & I_{12}(\theta) \\ I_{21}(\theta) & I_{22}(\theta) \end{pmatrix},
$$

where $H_1(\theta)$ is $p \times 1$ and each of $Q_{11}(\theta)$ and $I_{11}(\theta)$ is $p \times p$. Let $I_{11\cdot 2}(\theta) = I_{11}(\theta) - I_{12}(\theta)I_{22}^{-1}(\theta)I_{21}(\theta)$. Serial numbers $p+1, \ldots, r$ will be used for indexing the rows and columns of $I_{22}^{-1}(\theta)$ and for $p+1 \leq i, j \leq r$, let $a^{ij}(\theta)$ represent the

(i,j)-th element of $I_{22}^{-1}(\theta)$. Also, let $L_1(\theta)$ be a $p \times p^2$ submatrix of $L(\theta)$ with its i-th row given by $(K_{i11}(\theta), \ldots, K_{i1p}(\theta), \ldots, K_{ip1}(\theta), \ldots, K_{ipp}(\theta))$, $1 \le i \le p$.

Let $\hat{\theta}_{20}$ be the maximum likelihood estimator of θ_2 given $\theta_1 = \theta_{10}$. Any function evaluated at $\theta_1 = \theta_{10}$, $\theta_2 = \hat{\theta}_{20}$ will be distinguished by the addition of a circumflex (ˆ). Similarly, any function evaluated at $\theta_1 = \theta_{10}$, $\theta_2 = \theta_2$ or at $\theta_1 = \theta_{1n}$, $\theta_2 = \theta_2$ will be distinguished by the addition of a horizontal bar (¯) or a tilde (˜) respectively. Thus, for $1 \le i, j, u \le r$, $\hat{K}_{iju} = K_{iju}(\theta_{10}, \hat{\theta}_{20})$, $\bar{K}_{iju} = K_{iju}(\theta_{10}, \theta_2)$, $\tilde{K}_{iju} = K_{iju}(\theta_{1n}, \theta_2)$, and so on.

For ease in reference, before concluding this section, we state some basic facts about the usual likelihood ratio statistic for testing the null hypothesis $H_0 : \theta_1 = \theta_{10}$. This is given by

$$S = 2\{l_X(\hat{\theta}_1, \hat{\theta}_2) - l_X(\theta_{10}, \hat{\theta}_{20})\}, \tag{2.1}$$

where $l_X(\theta) = l_X(\theta_1, \theta_2) = \sum_{s=1}^{n} \log f(X_s; \theta)$, and $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ is the unrestricted maximum likelihood estimator of θ. Following Hayakawa (1975) (see also Hayakawa (1977)), it can be seen that under contiguous alternatives $\theta_{1n} = \theta_{10} + n^{-1/2}d$,

$$S = \hat{H}'\hat{I}^{-1}\hat{H} + n^{-1/2}\left[\hat{H}'\hat{I}^{-1}\hat{Q}\hat{I}^{-1}\hat{H} + \frac{1}{3}\hat{H}'\hat{I}^{-1}\hat{L}\{(\hat{I}^{-1}\hat{H}) \otimes (\hat{I}^{-1}\hat{H})\}\right] + o(n^{-1/2}), \tag{2.2}$$

where $\otimes$ denotes Kronecker product. Incidentally, by the definition of $\hat{H}$, the last $r-p$ elements of $\hat{H}$ are zeros. If one considers the usual likelihood ratio test based on the critical region

$$S > z^2 \tag{2.3}$$

where z^2 is the upper α-point of a central chi-square variate with p degrees of freedom, then, as shown in Hayakawa (1975) (see also Harris and Peers (1980)), its power function, under contiguous alternatives $\theta_{1n} = \theta_{10} + n^{-1/2}d$, is given by

$$P(d, \theta_2) = P_0(d, \theta_2) + n^{-1/2}P_1(d, \theta_2) + o(n^{-1/2}), \tag{2.4}$$

where

$$P_0(d, \theta_2) = G_{p,\lambda}(z^2), \qquad P_1(d, \theta_2) = \sum_{j=0}^{2} m_j G_{p+2j,\lambda}(z^2), \tag{2.5}$$

$\lambda = d'\bar{I}_{11\cdot 2}d$, $G_{\nu,\lambda}(z^2)$ is the probability for a non-central chi-square variate, with degrees of freedom ν and non-centrality parameter λ, to exceed z^2,

$$m_0 = -\frac{1}{6}\Big[\sum\sum\sum\{(\bar{K}_{i,j,u} - \bar{K}_{iju})\bar{d}_i^*\bar{d}_j^*\bar{d}_u^* - (3\bar{K}_{iju} + 6\bar{K}_{i,ju})\bar{\mu}_{ij}\bar{d}_u^*\} \tag{2.6a}$$

$$-\sum_1\sum\sum(3\bar{K}_{iju}+3\bar{K}_{i,ju})d_i\bar{d}_j^*\bar{d}_u^*\Bigg],$$

(2.6b) $$m_1=-\frac{1}{6}\Bigg[\sum\sum\sum\{(\bar{K}_{iju}-2\bar{K}_{i,j,u})\bar{d}_i^*\bar{d}_j^*\bar{d}_u^* + (3\bar{K}_{iju}+6\bar{K}_{i,ju})\bar{\mu}_{ij}\bar{d}_u^*\} + \sum_1\sum\sum(3\bar{K}_{iju}+3\bar{K}_{i,ju})d_i\bar{d}_j^*\bar{d}_u^*\Bigg],$$

(2.6c) $$m_2=-\frac{1}{6}\sum\sum\sum\bar{K}_{i,j,u}\bar{d}_i^*\bar{d}_j^*\bar{d}_u^*.$$

In (2.6a–c) and elsewhere in this paper, $\sum\sum\sum$ denotes summation over i, j, u in the range $1\le i,j,u\le r$ and $\sum_1\sum\sum$ denotes summation over i, j, u in the range $1\le i\le p$, $1\le j,u\le r$. Furthermore, for $1\le i,j\le r$, $\bar{d}_i^*=d_i^*(\theta_{10},\theta_2)$, $\bar{\mu}_{ij}=\mu_{ij}(\theta_{10},\theta_2)$, where $d_i^*(\theta)$ is the i-th element and $\mu_{ij}(\theta)$ is the (i,j)-th element of

$$d^*(\theta)=\begin{pmatrix}-\mathcal{G}_p\\ I_{22}^{-1}(\theta)I_{21}(\theta)\end{pmatrix}d \quad\text{and}\quad \mu(\theta)=\begin{pmatrix}0 & 0\\ 0 & I_{22}^{-1}(\theta)\end{pmatrix}$$

respectively, and $\mathcal{G}_p$ is the $p\times p$ identity matrix.

Taking $d=0$ in (2.4), it is easily seen that the likelihood ratio test given by (2.3) has size $\alpha+o(n^{-1/2})$. In the special case, where θ_1 and θ_2 are both one-dimensional and global parametric orthogonality holds, a tedious algebra shows that (2.4) agrees with the corresponding expression given in Mukerjee (1992a).

3. A new test

Let $g^*(\theta)=(g_1^*(\theta),\ldots,g_p^*(\theta))'$, where for $1\le s\le p$,

(3.1) $$g_s^*(\theta)=\sum\sum\sum\sum_{i,j,u,q=p+1}^{r}a^{ij}(\theta)a^{uq}(\theta)I_{(qs)}(\theta)\left\{\frac{1}{2}K_{iju}(\theta)+K_{i,ju}(\theta)\right\} -\sum\sum_{i,j=p+1}^{r}a^{ij}(\theta)\left\{\frac{1}{2}K_{ijs}(\theta)+K_{i,js}(\theta)\right\}.$$

Define

(3.2) $$g(\theta)=(g_1(\theta),\ldots,g_p(\theta))'=I_{11\cdot2}^{-1}(\theta)g^*(\theta).$$

Observe that (3.1) can be expressed in terms of the second-order biases of maximum likelihood estimators—see e.g., Cox and Snell (1968) and Cordeiro and McCullagh (1991). Furthermore, by (2.6a–c), the terms in m_0 and m_1 that are linear in d have coefficient vectors $\pm\bar{g}^*$, where $\bar{g}^*=g^*(\theta_{10},\theta_2)$. Therefore, $g^*(\theta)$ (and hence $g(\theta)$) can also be interpreted as an entity that accounts for a possible second-order local bias in the usual likelihood ratio test given by (2.3). As a modified version of the likelihood ratio statistic, we propose the statistic

(3.3) $$S^*=S+2n^{-1/2}\hat{H}_1'\hat{g},$$

where S is given by (2.1). We consider a test based on the critical region

$$S^* > z^2. \tag{3.4}$$

The power function, up to the second order and under contiguous alternatives, of the test given by (3.4) will be obtained by inverting the approximate characteristic function of S^*. This approach, different from that in Mukerjee (1992*a*) but similar to that considered in Hayakawa (1975), is convenient since one can utilize the findings in the latter paper. The formal computations in this section can be justified along the line of Chandra and Ghosh (1980).

By the definition of $\hat{\theta}_{20}$, one obtains $\hat{H}_2 = 0$. Hence a Taylor's expansion about (θ_{1n}, θ_2) and some subsequent simplification yield

$$\hat{\theta}_{20} = \theta_2 + n^{-1/2}\bar{I}_{22}^{-1}(\tilde{H}_2 + \bar{I}_{21}d) + o(n^{-1/2}). \tag{3.5}$$

Similarly, by (2.1) and (3.5),

$$S = (\tilde{H}_1 - \bar{I}_{12}\bar{I}_{22}^{-1}\tilde{H}_2 + \bar{I}_{11\cdot 2}d)'\bar{I}_{11\cdot 2}^{-1}(\tilde{H}_1 - \bar{I}_{12}\bar{I}_{22}^{-1}\tilde{H}_2 + \bar{I}_{11\cdot 2}d) + o(1), \tag{3.6a}$$

$$\hat{H}_1'\bar{g} = (\tilde{H}_1 - \bar{I}_{12}\bar{I}_{22}^{-1}\tilde{H}_2 + \bar{I}_{11\cdot 2}d)'\bar{g} + o(1). \tag{3.6b}$$

Hence writing $\xi = (-1)^{1/2}t$, and E_n for E_{θ_{1n},θ_2}, by (3.3), (3.6a, b), the approximate characteristic function of S^* under (θ_{1n}, θ_2) is given by

$$\begin{aligned} E_n\{\exp(\xi S^*)\} = E_n\{\exp(\xi S)\} &+ 2\xi n^{-1/2}E\{(\bar{g}'Y)\exp(\xi Y'\bar{I}_{11\cdot 2}^{-1}Y)\} \\ &+ o(n^{-1/2}), \end{aligned} \tag{3.7}$$

where Y is p-variate normal with mean vector $\bar{I}_{11\cdot 2}d$ and dispersion matrix $\bar{I}_{11\cdot 2}$. This step is explained by the fact that under (θ_{1n}, θ_2), up to the first order of approximation, $\tilde{H}_1 - \bar{I}_{12}\bar{I}_{22}^{-1}\tilde{H}_2$ is p-variate normal with null mean vector and dispersion matrix $\bar{I}_{11\cdot 2}$. As reported in Hayakawa (1975),

$$E_n\{\exp(\xi S)\} = \eta(p, \lambda, \xi) + n^{-1/2}\sum_{j=0}^{2} m_j\eta(p + 2j, \lambda, \xi) + o(n^{-1/2}), \tag{3.8}$$

where m_0, m_1, m_2 are as in (2.6a–c), $\lambda = d'\bar{I}_{11\cdot 2}d$ as before, and $\eta(\nu, \lambda, \xi)$ is the characteristic function of a non-central chi-square variate with degrees of freedom ν and non-centrality parameter λ. It can also be seen that

$$2\xi E\{(\bar{g}'Y)\exp(\xi Y'\bar{I}_{11\cdot 2}^{-1}Y)\} = (\bar{g}'\bar{I}_{11\cdot 2}d)\{\eta(p + 2, \lambda, \xi) - \eta(p, \lambda, \xi)\}. \tag{3.9}$$

By (3.1), (3.2) and the definitions of $\bar{d}_i^*$, $\bar{\mu}_{ij}$ $(1 \le i, j \le p)$, after some simplification,

$$\bar{g}'\bar{I}_{11\cdot 2}d = (\bar{g}^*)'d = \sum\sum\sum\left(\frac{1}{2}\bar{K}_{iju} + \bar{K}_{i,ju}\right)\bar{\mu}_{ij}\bar{d}_u^*.$$

Hence by (2.6a–c), (3.7)–(3.9),

$$E_n\{\exp(\xi S^*)\} = \eta(p,\lambda,\xi) + n^{-1/2}\sum_{j=0}^{2} m_j^*\eta(p+2j,\lambda,\xi) + o(n^{-1/2}), \tag{3.10}$$

where

$$m_0^* = m_0 - \bar{g}'\bar{I}_{11\cdot 2}d = -\frac{1}{6}\Big[\sum\sum\sum(\bar{K}_{i,j,u} - \bar{K}_{iju})\bar{d}_i^*\bar{d}_j^*\bar{d}_u^* - \sum_1\sum\sum(3\bar{K}_{iju} + 3\bar{K}_{i,ju})d_i\bar{d}_j^*\bar{d}_u^*\Big], \tag{3.11a}$$

$$m_1^* = m_1 + \bar{g}'\bar{I}_{11\cdot 2}d = -\frac{1}{6}\Big[\sum\sum\sum(\bar{K}_{iju} - 2\bar{K}_{i,j,u})\bar{d}_i^*\bar{d}_j^*\bar{d}_u^* + \sum_1\sum\sum(3\bar{K}_{iju} + 3\bar{K}_{i,ju})d_i\bar{d}_j^*\bar{d}_u^*\Big], \tag{3.11b}$$

$$m_2^* = m_2 = -\frac{1}{6}\sum\sum\sum \bar{K}_{i,j,u}\bar{d}_i^*\bar{d}_j^*\bar{d}_u^*. \tag{3.11c}$$

As in Hayakawa (1975), in consideration of (3.10), the power function of the test in (3.4), under contiguous alternatives $\theta_{1n} = \theta_{10} + n^{-1/2}d$, is given by

$$P^*(d,\theta_2) = P_0^*(d,\theta_2) + n^{-1/2}P_1^*(d,\theta_2) + o(n^{-1/2}), \tag{3.12}$$

where

$$P_0^*(d,\theta_2) = G_{p,\lambda}(z^2), \qquad P_1^*(d,\theta_2) = \sum_{j=0}^{2} m_j^* G_{p+2j,\lambda}(z^2), \tag{3.13}$$

m_0^*, m_1^*, m_2^* being given by (3.11a–c).

Taking $d = 0$ in (3.12), it can be seen that the test based on S^* as given by (3.4), like that based on S as given by (2.3), has size $\alpha + o(n^{-1/2})$. Hence it would be meaningful to compare these two tests in terms of power up to $o(n^{-1/2})$. From (2.5), (3.13) and (3.11a–c), note that $P_0(d,\theta_2) = P_0^*(d,\theta_2)$ identically in d and θ_2 while $P_1(d,\theta_2)$ and $P_1^*(d,\theta_2)$ are not identical unless $\bar{g} = 0$ identically in θ_2. Thus, as one can anticipate, the two tests have identical first-order power while, unless $\bar{g} = 0$ identically in θ_2, they can be discriminated in terms of second-order power.

Our next result relates to the superiority of the test based on S^* to the usual likelihood ratio test based on S with regard to second-order local maximinity. For each fixed Δ (> 0) and each fixed θ_2, let

$$Z(\Delta,\theta_2) = \min P_1(d,\theta_2), \qquad Z^*(\Delta,\theta_2) = \min P_1^*(d,\theta_2), \tag{3.14}$$

the minimum in either case being over d such that $d'\bar{I}_{11\cdot 2}d = \Delta$. Thus for each fixed θ_2 and Δ, minimization is done in (3.14) along spheres, centred at $d = 0$, with $\bar{I}_{11\cdot 2}$ used as a Riemannian metric (compare, Amari ((1985), Chapter 2)). We recall that $\bar{g}$ is a function of θ_2 and write $\bar{g} = \bar{g}(\theta_2)$. Then the following lemma, proved in the Appendix, holds.

LEMMA 3.1. *Suppose $\bar{g}(\theta_2)$ is not identically (in θ_2) equal to the null vector and consider a fixed θ_2 such that $\bar{g}(\theta_2) \neq 0$. Then there exists a positive $\Delta_0 = \Delta_0(\theta_2)$, depending on θ_2, such that $Z^*(\Delta, \theta_2) > Z(\Delta, \theta_2)$, whenever $0 < \Delta < \Delta_0(\theta_2)$.*

From (2.5), (3.13), note also that for fixed Δ (> 0) and θ_2 and for each d satisfying $d'\bar{I}_{11\cdot 2}d = \Delta$, the relation $P_0(d, \theta_2) = P_0^*(d, \theta_2) = G_{p,\Delta}(z^2)$ holds. Hence by (2.4), (3.12), (3.14) and Lemma 3.1, it is clear that, under the criterion of local maximinity and up to the second order of comparison, the test based on S^* is always at least as good as the usual likelihood ratio test and better whenever $\bar{g} = \bar{g}(\theta_2)$ is not identically (in θ_2) equal to the null vector. This result is rather strong and does not require any assumption regarding global parametric orthogonality. There are many models of practical importance where $\bar{g}$ is not identically equal to the null vector. An illustrative example is presented in the next section.

Remark. In consideration of (3.1)–(3.3), the computation of the statistic S^* calls for evaluation of expectations like $I_{(ij)}(\theta)$, $K_{iju}(\theta)$, $K_{i,ju}(\theta)$. For some models, it may be difficult to obtain these expectations analytically. Essentially in the spirit of Cox and Reid (1987), in such situations one may consider an alternative but equivalent statistic as indicated below. Let $J(\theta) = ((J_{(ij)}(\theta)))$ be the $r \times r$ observed information matrix at θ, where $J_{(ij)}(\theta) = -n^{-1} \sum_{s=1}^{n} D_i D_j \log f(X_s; \theta)$, $1 \leq i, j \leq r$. Similarly, for $1 \leq i, j, u \leq r$, let

$$k_{iju}(\theta) = n^{-1} \sum_{s=1}^{n} D_i D_j D_u \log f(X_s; \theta),$$

$$k_{i,ju}(\theta) = n^{-1} \sum_{s=1}^{n} \{D_i \log f(X_s; \theta)\}\{D_j D_u \log f(X_s; \theta)\}$$

be the 'observed' entities corresponding to $K_{iju}(\theta)$, $K_{i,ju}(\theta)$ respectively. Also, let $\hat{b}^{ij}$ ($p + 1 \leq i, j \leq r$) and $\hat{J}_{11\cdot 2}$ be defined with reference to $\hat{J}$ exactly as $\hat{a}^{ij}$ ($p + 1 \leq i, j \leq r$) and $\hat{I}_{11\cdot 2}$ are respectively defined with reference to $\hat{I}$. Then, in analogy with (3.1)–(3.3), one can consider the statistic $S^{**} = S + 2n^{-1/2}\hat{H}_1'\hat{\beta}$, where $\hat{\beta} = \hat{J}_{11\cdot 2}^{-1}\hat{\beta}^*$, and $\hat{\beta}^* = (\hat{\beta}_1^*, \ldots, \hat{\beta}_p^*)'$, with

$$\hat{\beta}_s^* = \sum\sum\sum\sum_{i,j,u,q=p+1}^{r} \hat{b}^{ij}\hat{b}^{uq}\hat{J}_{(qs)}\left(\frac{1}{2}\hat{k}_{iju} + \hat{k}_{i,ju}\right) - \sum\sum_{i,j=p+1}^{r} \hat{b}^{ij}\left(\frac{1}{2}\hat{k}_{ijs} + \hat{k}_{i,js}\right),$$

$1 \le s \le p$. It is not hard to see that $S^{**} = S^* + o(n^{-1/2})$. As such, up to $o(n^{-1/2})$, the power function, under contiguous alternatives, of a test based on S^{**} will be identical with that of the test based on S^*.

4. Parametric orthogonality and connexion with conditional likelihood

In this section, we consider the situation where global parametric orthogonality holds, that is, $I_{12}(\theta) = 0$, $I_{21}(\theta) = 0$, identically in θ. Then for $1 \le i \le r$, $p+1 \le j \le r$, $1 \le s \le p$, the regularity condition $K_{ijs}(\theta) + K_{i,js}(\theta) = 0$ holds. Hence by (3.1), (3.2),

$$g(\theta) = I_{11}^{-1}(\theta) g^*(\theta), \tag{4.1}$$

where $g^*(\theta) = (g_1^*(\theta), \ldots, g_p^*(\theta))'$, with

$$g_s^*(\theta) = \frac{1}{2} \sum\sum_{i,j=p+1}^{r} a^{ij}(\theta) K_{ijs}(\theta), \qquad 1 \le s \le p. \tag{4.2}$$

Therefore, by (3.3),

$$S^* = S + 2n^{-1/2} \hat{H}_1' \hat{I}_{11}^{-1} \hat{g}^*. \tag{4.3}$$

Under global parametric orthogonality, $\bar{d}_i^* = -d_i$ if $1 \le i \le p$, and $= 0$ otherwise. As a consequence, by (3.11a–c) and standard regularity conditions,

$$m_0^* = m_0 - \bar{g}' \bar{I}_{11} d = \frac{1}{6} \sum\sum\sum_{i,j,u=1}^{p} \bar{K}_{iju} d_i d_j d_u, \tag{4.4a}$$

$$m_1^* = m_1 + \bar{g}' \bar{I}_{11} d = \frac{1}{2} \sum\sum\sum_{i,j,u=1}^{p} \bar{K}_{i,ju} d_i d_j d_u, \tag{4.4b}$$

$$m_2^* = m_2 = \frac{1}{6} \sum\sum\sum_{i,j,u=1}^{p} \bar{K}_{i,j,u} d_i d_j d_u. \tag{4.4c}$$

Also, now λ, as in (2.5) and (3.13), equals $d' \bar{I}_{11} d$. Hence by (3.12), (3.13), (4.4a–c) and some results in Peers (1971), as corrected in Hayakawa (1975), it follows that for each fixed θ_2, under global parametric orthogonality and up to $o(n^{-1/2})$, the power function, under contiguous alternatives, of the test based on S^* is identical (in d) with that of a likelihood ratio test with known nuisance parameter. It is also evident that this does not hold with the usual likelihood ratio test based on S unless $\bar{g} = 0$ identically in θ_2. In a sense, this implies that under global parametric orthogonality, the use of S^* rather than S neutralizes the effect of an unknown nuisance parameter. To summarize, under global parametric orthogonality and up to the second order of comparison, the test based on S^* will be superior to that based on S not only in terms of local maximinity but also with regard to proximity to the power function attainable by a likelihood ratio test with known θ_2 whenever $\bar{g}$ is not null identically in θ_2.

Example 1. Let $f(x;\theta)$ represent the v-variate normal model with unknown mean vector $\tau = (\tau_1, \ldots, \tau_v)'$ and an unknown dispersion matrix $W = ((w_{ij}))$, where $\tau \in \mathcal{R}^v$, W is positive definite, $\theta_1 = (w_{11}, \ldots, w_{1v}, w_{22}, \ldots, w_{2v}, \ldots, w_{vv})'$, $\theta_2 = \tau$. Consider $H_0 : \theta_1 = \theta_{10}$, where $\theta_{10} = (w_{110}, \ldots, w_{1v0}, w_{220}, \ldots, w_{2v0}, \ldots, w_{vv0})'$ and the matrix $W_0 = ((w_{ij0}))$ is positive definite. It is easily seen that here global parametric orthogonality holds and that $I_{22}^{-1}(\theta) = W$. Furthermore, for $1 \le i, j, u, s \le v$, $u \le s$,

$$E_\theta\{\partial^3 \log f(X;\theta)/\partial\tau_i\partial\tau_j\partial w_{us}\} = -\partial w^{ij}/\partial w_{us},$$

where $W^{-1} = ((w^{ij}))$. Since for $1 \le u \le s \le v$, $\sum\sum_{i,j=1}^{v} w_{ij}(-\partial w^{ij}/\partial w_{us})$ equals w^{uu} for $u = s$, and $2w^{us}$ for $u < s$, and since the diagonal elements of W_0^{-1} are all positive, it follows from (4.1), (4.2) that in this example, $\bar{g} \ne 0$ for every θ_2. Therefore, the test given by S^* will be superior to that given by S in the senses described above.

It will now be seen how, under global parametric orthogonality, the statistic S^* arises from the principle of conditional likelihood. As defined in Cox and Reid (1987), under global parametric orthogonality, the conditional likelihood ratio statistic is given by

$$S_{\text{cond}} = 2\{l_{X,\text{cond}}(\theta_{1,\text{cond}}) - l_{X,\text{cond}}(\theta_{10})\}, \tag{4.5}$$

where

$$l_{X,\text{cond}}(\theta_1) = l_X(\theta_1, \hat{\theta}_2(\theta_1)) - \frac{1}{2}\log\det\{nJ_{22}(\theta_1, \hat{\theta}_2(\theta_1))\},$$
$$l_{X,\text{cond}}(\theta_{1,\text{cond}}) = \sup_{\theta_1} l_{X,\text{cond}}(\theta_1),$$

$\hat{\theta}_2(\theta_1)$ is the maximum likelihood estimator of θ_2 given θ_1, and $J_{22}(\theta)$ is the principal submatrix of $J(\theta)$ given by the last $r-p$ rows and columns of $J(\theta)$. After a considerable algebra (compare, McCullagh ((1987), Chapter 7)) one obtains

$$\begin{aligned}\theta_{1,\text{cond}} = \theta_{10} &+ n^{-1/2}\hat{I}_{11}^{-1}\hat{H}_1 \\ &+ n^{-1}\Big[\hat{I}_{11}^{-1}\hat{g}^* + \hat{I}_{11}^{-1}\hat{Q}_{11}\hat{I}_{11}^{-1}\hat{H}_1 \\ &\qquad + \frac{1}{2}\hat{I}_{11}^{-1}\hat{L}_1\{(\hat{I}_{11}^{-1}\hat{H}_1)\otimes(\hat{I}_{11}^{-1}\hat{H}_1)\}\Big] + o(n^{-1}),\end{aligned} \tag{4.6}$$

where $\hat{g}^*$ corresponds to (4.2) and $\hat{Q}_{11}$, $\hat{L}_1$ are as defined in Section 2. In the special case $p = 1$, $r = 2$, it can be seen that (4.6) is in agreement with relations (2.1a) and (2.3) of Mukerjee and Chandra (1991). From (4.3), (4.5), (4.6) and a reduced version of (2.2) under global parametric orthogonality, it can be seen that

$$S_{\text{cond}} = S^* + o(n^{-1/2}). \tag{4.7}$$

Hence, under global parametric orthogonality, the statistic S^* can be motivated by and obtained from the principle of conditional likelihood. In fact, it represents an expansion, up to $o(n^{-1/2})$, for the conditional likelihood ratio statistic. By the results proved in Section 3 and earlier in this section and (4.7), it also follows that, under global parametric orthogonality and up to the second order of comparison, the conditional likelihood ratio test will be at least as good as the usual likelihood ratio test with regard to (a) local maximinity, and (b) proximity to the power attainable by a likelihood ratio test with known nuisance parameter, and will, in fact, be better unless $\bar{g} = 0$ identically in θ_2. This generalizes the earlier results in Mukerjee (1992*a*) who considered the special case $p = 1$. It is interesting to note that in the absence of global parametric orthogonality, conditional likelihood, as in Cox and Reid (1987), is not well-defined but the test given by S^* continues to remain well-defined and possess the desirable property mentioned in (a) above. In consideration of the above, it appears that the test based on S^* can be regarded as a meaningful extension of the conditional likelihood ratio test to a general multiparameter set-up where global parametric orthogonality may or may not hold.

Under global parametric orthogonality, starting from the conditional likelihood $l_{X,\text{cond}}(\theta_1)$, it is also possible to define conditional versions of score and Wald's statistics. Mukerjee (1992*b*) studied the power properties of such versions for the case $p = 1$. It should be possible to extend these conditional versions to a general multiparameter set-up by considering modifications similar to (3.3) and then to study their power properties using the techniques employed here.

5. Relation with adjusted likelihood

We now return to the general set-up where no assumption is made regarding parametric orthogonality and indicate how the statistic S^* can be motivated also from a simplified version of the principle of adjusted likelihood considered by McCullagh and Tibshirani (1990). This will be done for the case $p = 1$ since, as noted in McCullagh and Tibshirani (1990), the adjusted likelihood (or its present simplified version) may not be well-defined for $p > 1$. Considering one-dimensional θ_1 and proceeding along the line of McCullagh and Tibshirani (1990), the adjusted likelihood is defined as

$$l_{X,\text{adj}}(\theta_1) = \int^{\theta_1} U_1(t)dt \tag{5.1a}$$

where

$$U_1(\theta_1) = U(\theta_1) - C(\theta_1), \tag{5.1b}$$

$$C(\theta_1) = E_{\theta_1,\hat{\theta}_2(\theta_1)}\{U(\theta_1)\}, \qquad U(\theta_1) = D_1 l_X(\theta_1, \hat{\theta}_2(\theta_1)). \tag{5.1c}$$

Note that, like McCullagh and Tibshirani (1990), we are correcting the mean of the score function but, unlike them, we are not adjusting its variance; see also Conniffe (1990) in this context. An adjusted likelihood ratio statistic may be defined as

$$S_{\text{adj}} = 2\{l_{X,\text{adj}}(\theta_{1,\text{adj}}) - l_{X,\text{adj}}(\theta_{10})\}, \tag{5.2}$$

where $\theta_{1,\mathrm{adj}}$ satisfies $l_{X,\mathrm{adj}}(\theta_{1,\mathrm{adj}}) = \sup_{\theta_1} l_{X,\mathrm{adj}}(\theta_1)$.

For each fixed θ, considering an expansion for $U(\theta_1)$ over a set with P_θ-probability $1 + o(n^{-1/2})$, it follows from relation (7) in McCullagh and Tibshirani (1990) and standard regularity conditions that

$$E_\theta\{U(\theta_1)\} = -g_1^*(\theta) + o(n^{-1/2}), \tag{5.3}$$

where $g_1^*(\theta)$ is given by (3.1). Hence, after some algebra, it can be seen that $\theta_{1,\mathrm{adj}} = \hat{\theta}_1 + o(n^{-1/2})$, where $\hat{\theta}_1$ is the (unrestricted) maximum likelihood estimator of θ_1, and that

$$\begin{aligned} S_{\mathrm{adj}} &= 2\{l_{X,\mathrm{adj}}(\hat{\theta}_1) - l_{X,\mathrm{adj}}(\theta_{10})\} + o(n^{-1/2}) \\ &= S + 2n^{-1/2}\hat{H}_{(1)}\hat{I}_{11\cdot 2}^{-1}\hat{g}_1^* + o(n^{-1/2}), \end{aligned}$$

by (5.1a–c), (5.2), (5.3). Comparing with (3.2), (3.3), it follows that $S_{\mathrm{adj}} = S^* + o(n^{-1/2})$, for $p = 1$. Thus for $p = 1$, even under the absence of parametric orthogonality, the statistic S^* can be obtained from and motivated by consideration of adjusted likelihood along the line of McCullagh and Tibshirani (1990).

While concluding, we remark that in view of the discussion in the last section and the findings in Mukerjee and Chandra (1991) for the simple case $p = 1$, $r = 2$, the statistic S^* is anticipated to admit a Bartlett-type adjustment. If such an adjustment is possible then the resulting statistic will be equivalent to S^*, up to $o(n^{-1/2})$ and, therefore, will continue to enjoy the desirable second-order properties of S^* proved earlier. A detailed study on the anticipated Bartlett-type adjustment calls for third-order calculations and this will be taken up in future. It is believed that a Bayesian route, along the line of Ghosh and Mukerjee (1992), will be helpful in this regard. It will also be interesting in future to make a small sample comparison of S and S^*, by simulation studies with reference to some problems of practical interest, to show that the modification proposed in (3.3) may be substantial even in small samples; see Mukerjee (1992*b*) for such a numerical study with $p = 1$.

Acknowledgements

This work was supported by a grant from the Centre for Management and Development Studies, Indian Institute of Management Calcutta. Thanks are also due to a referee for very constructive suggestions.

Appendix

PROOF OF LEMMA 3.1. Recall that $\bar{I}_{11\cdot 2}$ is a function of θ_2 and write $\bar{I}_{11\cdot 2} = \bar{I}_{11\cdot 2}(\theta_2)$. For $\Delta > 0$ and $d = (d_1, \ldots, d_p)'$ satisfying $d'\bar{I}_{11\cdot 2}(\theta_2)d = \Delta$, note that

$$|d_i| \leq \{\Delta/\phi(\theta_2)\}^{1/2}, \qquad 1 \leq i \leq p, \tag{A.1}$$

where $\phi(\theta_2)$ (> 0) is the smallest eigenvalue of $\bar{I}_{11\cdot 2}(\theta_2)$. By (2.6a–c), (3.11a–c),

$$m_s^* = \sum\sum\sum B_{iju}^{(s)} d_i d_j d_u \qquad (s = 0, 1, 2),$$
$$m_0 = \sum\sum\sum B_{iju}^{(0)} d_i d_j d_u + \{\bar{g}(\theta_2)\}'\bar{I}_{11\cdot 2}(\theta_2)d,$$
$$m_1 = \sum\sum\sum B_{iju}^{(1)} d_i d_j d_u - \{\bar{g}(\theta_2)\}'\bar{I}_{11\cdot 2}(\theta_2)d,$$
$$m_2 = \sum\sum\sum B_{iju}^{(2)} d_i d_j d_u,$$

where for each i, j, u ($1 \le i, j, u \le p$) and s, $B_{iju}^{(s)} = B_{iju}^{(s)}(\theta_2)$ may involve θ_2 but not d. Hence by (2.5), (A.1), for each d satisfying $d'\bar{I}_{11\cdot 2}(\theta_2)d = \Delta$,

$$\text{(A.2)}\quad P_1(d,\theta_2) = \sum_{s=0}^{2}\sum\sum\sum B_{iju}^{(s)} d_i d_j d_u G_{p+2s,\Delta}(z^2) - \{\bar{g}(\theta_2)\}'\bar{I}_{11\cdot 2}(\theta_2)d\{G_{p+2,\Delta}(z^2) - G_{p,\Delta}(z^2)\}$$
$$\le B(\Delta,\theta_2)\Delta^{3/2} - \{\bar{g}(\theta_2)\}'\bar{I}_{11\cdot 2}(\theta_2)d\{G_{p+2,\Delta}(z^2) - G_{p,\Delta}(z^2)\},$$

where $B(\Delta,\theta_2) = \sum\sum\sum|\sum_{s=0}^{2} B_{iju}^{(s)} G_{p+2s,\Delta}(z^2)|\{\phi(\theta_2)\}^{-3/2}$. In a similar manner, for each d satisfying $d'\bar{I}_{11\cdot 2}(\theta_2)d = \Delta$, we have $P_1^*(d,\theta_2) \ge -B(\Delta,\theta_2)\Delta^{3/2}$. Hence,

$$\text{(A.3)}\qquad Z^*(\Delta,\theta_2) \ge -B(\Delta,\theta_2)\Delta^{3/2}, \qquad \text{for} \quad \Delta > 0.$$

Recall that $\bar{g}(\theta_2)$ is non-null. Hence $\bar{I}_{11\cdot 2}(\theta_2)\bar{g}(\theta_2)$ is non-null and, without loss of generality, suppose its first element, say $\zeta(\theta_2)$, is positive (the proof is similar for a negative element). Then $d = \{\Delta/\zeta^*(\theta_2)\}^{1/2}(1, 0, \ldots, 0)'$, where $\zeta^*(\theta_2)$ is the $(1,1)$-th element of $\bar{I}_{11\cdot 2}(\theta_2)$, satisfies $d'\bar{I}_{11\cdot 2}(\theta_2)d = \Delta$, and by (A.2),

$$\text{(A.4)}\quad Z(\Delta,\theta_2) \le B(\Delta,\theta_2)\Delta^{3/2} - \zeta(\theta_2)\{\Delta/\zeta^*(\theta_2)\}^{1/2}\{G_{p+2,\Delta}(z^2) - G_{p,\Delta}(z^2)\}, \quad \text{for} \quad \Delta > 0.$$

From (A.3), (A.4), it is straightforward to complete the proof.

References

Amari, S. (1985). *Differential Geometrical Methods in Statistics*, Springer, New York.

Barndorff-Nielsen, O. E. (1986). Inference on full or partial parameters based on the standardized log-likelihood ratio, *Biometrika*, **73**, 307–322.

Chandra, T. K. and Ghosh, J. K. (1979). Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables, *Sankhyā Ser. A*, **41**, 22–47.

Chandra, T. K. and Ghosh, J. K. (1980). Valid asymptotic expansions for the likelihood ratio and other statistics under contiguous alternatives, *Sankhyā Ser. A*, **42**, 170–184.

Conniffe, D. (1990). Testing hypotheses with estimated scores, *Biometrika*, **77**, 97–106.

Cordeiro, G. M. and McCullagh, P. (1991). Bias correction in generalized linear models, *J. Roy. Statist. Soc. Ser. B*, **53**, 629–643.

Cox, D. R. (1988). Some aspects of conditional and asymptotic inference: A review, *Sankhyā Ser. A*, **50**, 314–337.

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

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals (with discussion), *J. Roy. Statist. Soc. Ser. B*, **30**, 248–275.

Ghosh, J. K. and Mukerjee, R. (1992). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests, *J. Roy. Statist. Soc. Ser. B*, **54**, 867–875.

Godambe, V. P. (1991). Orthogonality of estimating functions and nuisance parameters, *Biometrika*, **78**, 143–151.

Harris, P. and Peers, H. W. (1980). The local power of the efficient scores test statistic, *Biometrika*, **67**, 525–529.

Hayakawa, T. (1975). The likelihood ratio criterion for a composite hypothesis under a local alternative, *Biometrika*, **62**, 451–460.

Hayakawa, T. (1977). The likelihood ratio criterion and the asymptotic expansion of its distribution, *Ann. Inst. Statist. Math.*, **29**, 359–378.

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

McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman and Hall, London.

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

Mukerjee, R. (1992*a*). Comparison between the conditional likelihood ratio and the usual likelihood ratio tests, *J. Roy. Statist. Soc. Ser. B*, **54**, 189–194.

Mukerjee, R. (1992*b*). Conditional likelihood and power: Higher order asymptotics, *Proc. Roy. Soc. London Ser. A*, **438**, 433–446.

Mukerjee, R. and Chandra, T. K. (1991). Bartlett-type adjustment for the conditional likelihood ratio statistic of Cox and Reid, *Biometrika*, **78**, 365–372.

Peers, H. W. (1971). Likelihood ratio and associated test criteria, *Biometrika*, **58**, 477–487.

Note added in proof: As in Section 5, the statistic S^* can be motivated also from modified likelihood as considered in the following paper:

Barndorff-Nielsen, O. E. (1994). Adjusted versions of profile likelihood and directed likelihood, and extended likelihood, *J. Roy. Statist. Soc. Ser. B*, **56** (to appear).