Journal of Statistical Planning and Inference 97 (2001) 45–55

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Test statistics arising from quasi likelihood: Bartlett adjustment and higher-order power

J.K. Ghosh[a,b], Rahul Mukerjee[c,*]

[a] Department of Statistics, Purdue University, West Lafayette, IN 47907, USA
[b] Indian Statistical Institute, 203 B.T. Road, Calcutta 700 035, India
[c] Indian Institute of Management, Post Box No. 16757, Calcutta 700 027, India

Abstract

With reference to the quasi-likelihood arising from an unbiased estimating function, we consider a large class of test statistics which includes the likelihood ratio, Rao's score and Wald's statistics in particular. We study Bartlett adjustability and third-order power in a possibly non-iid setting and provide explicit formulae. Since the relevant Bartlett identities may not hold while working with a quasi-likelihood, our results can differ from those based on the usual likelihood. The prospects regarding posterior Bartlett adjustability have been briefly indicated.

MSC: 62F05; 62E20

Keywords: Average power; Contiguous alternative; Likelihood ratio statistic; Maximinity; Rao's score statistic; Wald's statistic

1. Introduction

The study of Bartlett adjustability and higher-order power for test statistics based on the usual likelihood, under model specification, has received considerable attention in the literature — see e.g., Bickel and Ghosh (1990), Taniguchi (1991), Mukerjee (1993) and the references therein. The present paper investigates the corresponding problems with reference to the quasi-likelihood arising from an unbiased estimating function in a possibly non-iid setting. This is motivated by Barndorff-Nielsen's (1995) recent work on the first-order null distribution of the likelihood ratio (LR) statistic based on such a quasi-likelihood. Starting from an estimating function and the associated quasi-likelihood, we consider a large class of test statistics, which includes the LR, Rao's score and Wald's statistics, and characterize, via appropriate necessary and sufficient conditions; the Bartlett-adjustable members of this class. We also give an explicit formula for the third-order power function under contiguous alternatives and

* Corresponding author.
E-mail address: rmuk1@hotmail.com (R. Mukerjee).

0378-3758/01/$ - see front matter
PII: S0378-3758(00)00344-X

discuss its implications. The preliminaries and the main results are presented in Sections 2 and 3, respectively. The prospects regarding posterior Bartlett adjustability are briefly discussed in Section 4.

Since the relevant Bartlett identities may not hold while working with a quasi-likelihood, our results can differ from those based on the usual likelihood. For example, as seen later in Section 3, there can be situations where Rao's statistic is Bartlett adjustable but the LR statistic is not so. Some of our algebra has a similarity with that in Viraswami and Reid (1996a, b), who studied the null distributions of Rao's, Wald's and LR statistics under model misspecification. However, our findings are different from theirs. In particular, unlike them, we work within the framework of a large class of test statistics and our results involve computations under the null hypothesis as well as contiguous alternatives.

2. Preliminaries

Consider a collection $X^{(n)} = (X_1,\ldots,X_n)'$, $n \geqslant 1$, of possibly vector-valued random variables with a density involving an unknown parameter θ which belongs to an open subset of $\mathbb{R}^l$, Let $g_n(X^{(n)},\theta)$ be a smooth unbiased estimating function for θ and define

$$\psi_{in}(\theta) = E_\theta\left\{\frac{\mathrm{d}^i}{\mathrm{d}\theta^i}\, g_n(X^{(n)},\theta)\right\}\ (i=1,2,3), \quad V_n(\theta) = \mathrm{Var}_\theta\{g_n(X^{(n)},\theta)\}.$$

Following McCullagh (1991, Section 11.7), without loss of generality, it is supposed that

$$V_n(\theta) = -\psi_{1n}(\theta), \tag{2.1}$$

identically in θ. Let $\hat\theta$ be the estimator determined by $g_n(X^{(n)},\theta)$. As in Barndorff-Nielsen (1995), the quasi-likelihood, associated with $g_n(X^{(n)},\theta)$, is defined as

$$\mathscr{L}_n(X^{(n)},\theta) = \int_{\hat\theta}^{\theta} g_n(X^{(n)},\theta)\,\mathrm{d}\theta. \tag{2.2}$$

Considering the null hypothesis $\mathrm{H}_0 : \theta = \theta_0$ against $\theta \neq \theta_0$, the LR, Rao's and Wald's statistics arising from (2.2) are given, respectively, by

$$T^{\mathrm{LR}} = 2\{\mathscr{L}_n(X^{(n)},\hat\theta) - \mathscr{L}_n(X^{(n)},\theta_0)\},$$
$$T^{\mathrm{Rao}} = \{g_n(X^{(n)},\theta_0)\}^2/V_n(\theta_0), \quad T^{\mathrm{Wald}} = V_n(\hat\theta)(\hat\theta-\theta_0)^2, \tag{2.3}$$

cf. Rao (1973, p. 417). In what follows, we shall treat the statistics in (2.3) as members of a large class and study their properties.

We work under assumptions similar to those in Taniguchi (1991) — see also Chandra and Ghosh (1980). In particular, it is assumed that, for an appropriate sequence $\{c_n\}$

satisfying $c_n \to \infty$ as $n \to \infty$, the cumulants, up to fourth order, of

$$Z_1(\theta) = c_n^{-1} g_n(X^{(n)}, \theta), \quad Z_i(\theta) = c_n^{-1} \left\{ \frac{\mathrm{d}^{i-1}}{\mathrm{d}\theta^{\,i-1}} g_n(X^{(n)}, \theta) - \psi_{i-1\,n}(\theta) \right\} \ (i = 2, 3) \tag{2.4}$$

possess asymptotic expansions of the form

$$\mathrm{cum}_\theta\{Z_i(\theta), Z_j(\theta)\} = k_{ij}^{(1)}(\theta) + c_n^{-2} k_{ij}^{(2)}(\theta) + \mathrm{o}(c_n^{-2}), \tag{2.5a}$$

$$\mathrm{cum}_\theta\{Z_i(\theta), Z_j(\theta), Z_l(\theta)\} = c_n^{-1} k_{ijl}(\theta) + \mathrm{o}(c_n^{-2}), \tag{2.5b}$$

$$\mathrm{cum}_\theta\{Z_i(\theta), Z_j(\theta), Z_l(\theta), Z_u(\theta)\} = c_n^{-2} k_{ijlu}(\theta) + \mathrm{o}(c_n^{-2}), \tag{2.5c}$$

the fifth and subsequent cumulants being of order $\mathrm{o}(c_n^{-2})$. As in the iid case, the second-order cumulants have an expansion in powers of c_n^{-2} rather than c_n^{-1}; cf Chandra and Joshi (1983). The expansions in (2.5) as well as those in (2.7)–(2.9) below are assumed to be uniform over compact θ-subsets. The functions $k_{ij}^{(1)}(\theta)$, $k_{ij}^{(2)}(\theta)$, $k_{ijl}(\theta)$, $k_{ijlu}(\theta)$ are supposed to be smooth with functional forms free from n. Let

$$I(\theta) = k_{11}^{(1)}(\theta), \quad J(\theta) = k_{12}^{(1)}(\theta), \quad K(\theta) = k_{111}(\theta), \quad L(\theta) = k_{13}^{(1)}(\theta),$$

$$M(\theta) = k_{22}^{(1)}(\theta), \quad N(\theta) = k_{112}(\theta), \quad H(\theta) = k_{1111}(\theta), \quad \Delta(\theta) = k_{11}^{(2)}(\theta). \tag{2.6}$$

Note that by (2.1), (2.4) and (2.6),

$$c_n^{-2}\psi_{1n}(\theta) = -\mathrm{Var}_\theta\{Z_1(\theta)\} = -\{I(\theta) + c_n^{-2}\Delta(\theta)\} + \mathrm{o}(c_n^{-2}). \tag{2.7}$$

We assume that in analogy with what happens while working with the usual likelihood

$$c_n^{-2}\psi_{2n}(\theta) = m_2(\theta) + \mathrm{o}(c_n^{-1}), \quad c_n^{-2}\psi_{3n}(\theta) = m_3(\theta) + \mathrm{o}(1), \tag{2.8}$$

where $m_i(\theta)$, $i = 2, 3$, are smooth functions with functional forms free from n. It is also supposed that the first two derivatives of $c_n^{-2}\psi_{1n}(\theta)$ and the first derivative of $c_n^{-2}\psi_{2n}(\theta)$ can be approximated using (2.7) and the first relation in (2.8), respectively, i.e.,

$$c_n^{-2}\psi'_{1n}(\theta) = -\{I'(\theta) + c_n^{-2}\Delta'(\theta)\} + \mathrm{o}(c_n^{-2}), \tag{2.9}$$

and so on, where the primes denote differentiation with respect to θ. Write $Z_i = Z_i(\theta_0)$ $(i = 1, 2, 3)$, $m_i = m_i(\theta_0)$ $(i = 2, 3)$, $m_2' = m_2'(\theta_0)$, and, with reference to (2.5), let $I = I(\theta_0)$, $J = J(\theta_0)$, $K = K(\theta_0)$, $I' = I'(\theta_0)$, $I'' = I''(\theta_0)$, $J' = J'(\theta_0)$, and so on.

As mentioned earlier, for power studies we shall consider contiguous alternatives of the form $\theta_n = \theta_0 + c_n^{-1}h$, where h is free from n. Let $\mathscr{F}$ be a class of test statistics for $\mathrm{H}_0 : \theta = \theta_0$ such that every statistic T in $\mathscr{F}$ admits an expansion of the form

$$T = W_T^2 + \mathrm{o}(c_n^{-2}), \tag{2.10}$$

over a set with P_{θ_n}-probability $1 + \mathrm{o}(c_n^{-2})$, uniformly on compact subsets of h (this implies that $T = W_T^2 + \mathrm{o}_p(c_n^{-2})$), where

$$W_T = I^{-1/2} Z_1 + c_n^{-1}(v_1 Z_1 Z_2 + v_2 Z_1^2)$$
$$+ c_n^{-2}(y_1 Z_1 Z_2^2 + y_2 Z_1^2 Z_2 + y_3 Z_1^3 + y_4 Z_1^2 Z_3 + y_5 Z_1), \tag{2.11}$$

and $v_1, v_2, y_1, \ldots, y_5$ are constants free from n. The class $\mathscr{F}$, identical with that in Taniguchi (1991) and analogous to that in Chandra and Mukerjee (1985), is very rich and includes, in particular, the LR, Rao's and Wald's statistics the corresponding expressions for $v_1, v_2, y_1, \ldots, y_5$ being as shown in the appendix. Consideration of the square-root version W_T simplifies our algebra to some extent and helps in interpreting the conditions for Bartlett adjustability.

We shall require the first four approximate cumulants, under $\theta_n = \theta_0 + c_n^{-1}h$, of W_T. These are given by

$$Q_{1n} = I^{1/2}h + c_n^{-1}(f_1 + f_2h^2) + c_n^{-2}(f_3h + f_4h^3) + \mathrm{o}(c_n^{-2}), \tag{2.12a}$$

$$Q_{2n} = 1 + c_n^{-1} f_5h + c_n^{-2}(f_6 + f_7h^2) + \mathrm{o}(c_n^{-2}), \tag{2.12b}$$

$$Q_{3n} = c_n^{-1} f_8 + c_n^{-2} f_9h + \mathrm{o}(c_n^{-2}), \quad Q_{4n} = c_n^{-2} f_{10} + \mathrm{o}(c_n^{-2}), \tag{2.12c}$$

where the constants $f_1, \ldots, f_{10}$, free from n and h but dependent on $v_1, v_2, y_1, \ldots, y_5$, are as shown in the appendix. The expressions (2.12a–c) as well as the contents of the appendix follow from (2.3)–(2.9) and (2.11) proceeding along the line of Chandra and Joshi (1983) and Chandra and Mukerjee (1985). Unlike these authors, however, we work with a quasi likelihood and cannot in general employ the relevant Bartlett identities. As such, the quantities $f_1, \ldots, f_{10}$, shown in the appendix, are somewhat more involved than the corresponding expressions in Chandra and Mukerjee (1985). Incidentally, if one specializes to the usual likelihood and iid observations then the expressions in our appendix are in agreement with the findings of Chandra and Joshi (1983) and Chandra and Mukerjee (1985).

3. Main results

3.1. Third-order power

On the basis of any statistic T in $\mathscr{F}$, consider a critical region of the form $T > z^2 + c_n^{-2}s_T$, where z is the upper $\frac{1}{2}\alpha$-point of a standard normal variate and the constant s_T, free from n, is to be so determined that the test has size $\alpha + \mathrm{o}(c_n^{-2})$ $(0 < \alpha < 1)$. By (2.10) and (2.12), considering an Edgeworth expansion for the distribution of W_T under θ_n, after some simplification,

$$\begin{aligned} P_{\theta_n}(T > z^2 + c_n^{-2}s_T) &\\ = 1 - P_{\theta_n}(|W_T| \leqslant z + \tfrac{1}{2}c_n^{-2}z^{-1}s_T) + \mathrm{o}(c_n^{-2}) &\\ = 2 - \Phi(z - I^{1/2}h) - \Phi(z + I^{1/2}h) + c_n^{-1}\sum_{j=1}^{3}(-1)^j a_j(h)R_j(h) &\end{aligned}$$

$$+c_n^{-2}\left[\sum_{j=1}^{6}(-1)^j b_j(h)R_j(h)-\tfrac{1}{2}z^{-1}s_T\{\phi(z-I^{1/2}h)+\phi(z+I^{1/2}h)\}\right]$$

$$+\mathrm{o}(c_n^{-2}), \tag{3.1}$$

where ϕ and Φ are the density and distribution functions, respectively, of a standard normal variate,

$$a_1(h)=-(f_1+f_2h^2),\quad a_2(h)=\tfrac{1}{2}f_5h,\quad a_3(h)=-\tfrac{1}{6}f_8, \tag{3.2}$$

$$b_1(h)=-(f_3h+f_4h^3),\quad b_2(h)=\tfrac{1}{2}\{f_6+f_1^2+(f_7+2f_1f_2)h^2+f_2^2h^4\}, \tag{3.3a}$$

$$b_3(h)=-\{(\tfrac{1}{6}f_9+\tfrac{1}{2}f_1f_5)h+\tfrac{1}{2}f_2f_5h^3\}, \tag{3.3b}$$

$$b_4(h)=\tfrac{1}{24}f_{10}+\tfrac{1}{6}f_1f_8+(\tfrac{1}{8}f_5^2+\tfrac{1}{6}f_2f_8)h^2,\quad b_5(h)=-\tfrac{1}{12}f_5f_8h, \tag{3.3c}$$

$$b_6(h)=\tfrac{1}{72}f_8^2, \tag{3.3d}$$

$$R_j(h)=G_{j-1}(z-I^{1/2}h)\phi(z-I^{1/2}h)+(-1)^jG_{j-1}(z+I^{1/2}h)\phi(z+I^{1/2}h), \tag{3.4}$$

$G_{j-1}(.)$ being the Hermite polynomial of degree $j-1$. If one takes $h=0$ in (3.1) and employs the size condition, then one gets

$$s_T=2z\{b_2(0)G_1(z)+b_4(0)G_3(z)+b_6(0)G_5(z)\}.$$

Using this in (3.1), the third-order power function, under contiguous alternatives, of the test based on T is given by

$$P_{\theta_n}(T>z^2+c_n^{-2}s_T)=P_0(h)+c_n^{-1}P_1(h)+c_n^{-2}P_2(h)+\mathrm{o}(c_n^{-2}), \tag{3.5}$$

where

$$P_0(h)=2-\Phi(z-I^{1/2}h)-\Phi(z+I^{1/2}h),\quad P_1(h)=\sum_{j=1}^{3}(-1)^j a_j(h)R_j(h), \tag{3.6a}$$

$$P_2(h)=\sum_{j=1}^{6}(-1)^j b_j(h)R_j(h)-\{b_2(0)G_1(z)+b_4(0)G_3(z)$$
$$+b_6(0)G_5(z)\}\{\phi(z-I^{1/2}h)+\phi(z+I^{1/2}h)\}. \tag{3.6b}$$

For a meaningful comparison of the statistics in $\mathscr{F}$ with regard to power, we now bring in the criteria of maximinity and average power. This is in the spirit of what one does while working with the usual likelihood — see e.g., Mukerjee (1992, 1993). Under the criterion of maximinity, it is possible to discriminate among the members of $\mathscr{F}$ at the second order of approximation. By (3.2), (3.4) and (3.6a), $P_1(h)=\phi(z)S(z)h+\mathrm{O}(h^3)$, where

$$S(z)=(f_5+2f_1I^{1/2}-f_8I^{1/2})z+\tfrac{1}{3}f_8I^{1/2}z^3. \tag{3.7}$$

Consider now any two members $T^{(1)}$ and $T^{(2)}$ of $\mathscr{F}$, the corresponding expressions for $S(z)$ being denoted by $S^{(1)}(z)$ and $S^{(2)}(z)$, respectively. Proceeding as in Mukerjee

(1992), for any given α, $T^{(1)}$ will be better than $T^{(2)}$ with respect to second-order local maximinity, in the sense of rendering a larger value of

$$\min\{P_0(h)+c_n^{-1}P_1(h), P_0(-h)+c_n^{-1}P_1(-h)\}$$

for small $|h|$, provided

$$|S^{(1)}(z)| < |S^{(2)}(z)|. \tag{3.8}$$

Turning to the criterion of average power, we note that $P_0(h)$ and $P_2(h)$ are even functions while $P_1(h)$ is an odd function. Hence, as with maximinity, considering alternatives that are equidistant from θ_0, it follows from (3.5) that the third-order average power function is given by $P_0(h)+c_n^{-2}P_2(h)$. Now $P_0(h)$ is the same for all tests in $\mathscr{F}$ while by (3.3), (3.4) and (3.6b), $P_2(h)=\phi(z)U(z)h^2+\mathrm{O}(h^4)$, where

$$\begin{aligned} U(z) = {} & \{f_7+2f_1f_2+2I^{1/2}f_3-I(f_6+f_1^2)-I(\tfrac{1}{4}f_{10}+f_1f_8)\}G_1(z) \\ & +\{\tfrac{1}{4}f_5^2+\tfrac{1}{3}f_2f_8+I^{1/2}(\tfrac{1}{3}f_9+f_1f_5)-I(\tfrac{1}{4}f_{10}+f_1f_8)-\tfrac{5}{18}If_8^2\}G_3(z) \\ & +\{\tfrac{1}{6}I^{1/2}f_5f_8-\tfrac{5}{36}If_8^2\}G_5(z). \end{aligned} \tag{3.9}$$

Clearly, for given α, a statistic with a larger value of $U(z)$ will lead to a larger third-order local average power.

Unbiased estimating functions are too diverse to allow any general result on comparison of power in the present setup. However, in any specific setting as dictated by a given estimating function, it is possible to draw definite conclusions using (3.7)–(3.9). An illustrative example will be presented in Section 3.3.

3.2. Bartlett adjustability

By (2.10) and (2.12),

$$E_{\theta_0}(T)=1+c_n^{-2}(f_6+f_1^2)+\mathrm{o}(c_n^{-2}).$$

Considering an Edgeworth expansion for the distribution of W_T under θ_0, in analogy with (3.1), we get

$$\begin{aligned} & P_{\theta_0}[T/\{1+c_n^{-2}(f_6+f_1^2)\}\leqslant z^2] \\ & \quad = 2\Phi(z)-1-c_n^{-1}\sum_{j=1}^{3}(-1)^j a_j(0)R_j(0) \\ & \qquad - c_n^{-2}\left\{\sum_{j=1}^{6}(-1)^j b_j(0)R_j(0)-z\phi(z)(f_6+f_1^2)\right\}+\mathrm{o}(c_n^{-2}). \end{aligned}$$

Hence by (3.2)–(3.4),

$$\begin{aligned} P_{\theta_0}[T/\{1+c_n^{-2}(f_6+f_1^2)\}\leqslant z^2] = {} & 2\Phi(z)-1-c_n^{-2}\phi(z)\{\tfrac{1}{36}f_8^2G_5(z) \\ & +(\tfrac{1}{12}f_{10}+\tfrac{1}{3}f_1f_8)G_3(z)\}+\mathrm{o}(c_n^{-2}). \end{aligned}$$

The above equals $2\Phi(z) - 1 + o(c_n^{-2})$ for each z (> 0) if and only if

$$f_8 = 0, \quad f_{10} = 0 \tag{3.10}$$

which are the necessary and sufficient conditions for the Bartlett adjustability of T at θ_0. From (2.12c) and (3.10), it is interesting to note that T is Bart-adjustable at θ_0 if and only if the third and fourth cumulants of its square root version W_T are both of order $o(c_n^{-2})$ under θ_0.

Considering Rao's score statistic in particular, it can be seen from the appendix that $f_8 = I^{-3/2}K$ and $f_{10} = I^{-2}H$. Thus, Rao's statistic is Bartlett adjustable at every θ if and only if

$$K(\theta) = 0, \quad H(\theta) = 0 \tag{3.11}$$

identically in θ; incidentally, Rao's statistic has expectation unity under the null hypothesis and hence, under (3.11), even without any adjustment, its null distribution is chi-square with margin of error $o(c_n^{-2})$. From (3.10) and the appendix, it can also be seen that the LR statistic is Bartlett adjustable at every θ if and only if

$$K(\theta) + 3J(\theta) + m_2(\theta) = 0, \quad H(\theta) + 6N(\theta) + 3M(\theta) + 4L(\theta) + m_3(\theta) = 0, \tag{3.12}$$

identically in θ. If one works with the usual likelihood then the conditions in (3.12) are simply Bartlett identities so that the LR statistic becomes Bartlett adjustable. As the example in Section 3.3 reveals, with a quasi-likelihood, however, there is no guarantee that (3.12) will hold.

From the expressions for f_1, f_6, f_8 and f_{10}, as given in the appendix, one can check that conditions (3.10) for Bartlett adjustability as well as the Bartlett-adjustment factor $1+c_n^{-2}(f_6+f_1^2)$ arising under these conditions are in agreement with the findings in Taniguchi (1991) when one specializes to the usual likelihood.

3.3. An example

Let $X^{(n)}=(X_1,\ldots,X_n)'$, where $X_1,\ldots,X_n$ are independent, with $E_\theta(X_i)=\theta$, $\mathrm{Var}_\theta(X_i)=r_i\theta^2$, $1 \leqslant i \leqslant n$, the third and fourth cumulants being zero for each X_i. Here $\theta > 0$ and $r_1,\ldots,r_n$ are positive constants which are free from θ. Suppose

$$\bar{r}_n = (r_1 + \cdots + r_n)/n = 1 - n^{-1} + o(n^{-1}). \tag{3.13}$$

Consider the unbiased estimating function $g_n(X^{(n)},\theta) = n(\bar{X}_n - \theta)/(\bar{r}_n\theta^2)$, where $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Then (2.1) holds and with $c_n = n^{1/2}$,

$$\psi_{1n}(\theta) = -n/(\bar{r}_n\theta^2), \quad \psi_{2n}(\theta) = 4n/(\bar{r}_n\theta^3), \quad \psi_{3n}(\theta) = -18n/(\bar{r}_n\theta^4),$$

$$Z_1(\theta) = n^{1/2}(\bar{X}_n - \theta)/(\bar{r}_n\theta^2), \quad Z_2(\theta) = -(2/\theta)Z_1(\theta), \quad Z_3(\theta) = (6/\theta^2)Z_1(\theta).$$

Hence, by (2.5)–(2.8) and (3.13),

$$I(\theta) = \Delta(\theta) = 1/\theta^2, \quad J(\theta) = -2/\theta^3, \quad L(\theta) = 6/\theta^4, \quad M(\theta) = 4/\theta^4,$$

$$K(\theta) = N(\theta) = H(\theta) = 0, \quad m_2(\theta) = 4/\theta^3, \quad m_3(\theta) = -18/\theta^4. \tag{3.14}$$

By (3.11), (3.12), (3.14), Rao's statistic is Bartlett adjustable at each θ and the LR statistic is not so at any θ. From (3.10), (3.14) and the appendix, it can also be seen that Wald's statistic is not Bartlett adjustable at any θ.

We now consider $\mathrm{H}_0 : \theta = \theta_0$, where $\theta_0 = 1$. Then by (3.14),

$$I = 1, \quad I' = -2, \quad I'' = 6, \quad J = -2, \quad J' = 6, \quad K = K' = 0, \quad L = 6,$$

$$M = 4, \quad N = H = 0, \quad \Delta = 1,$$

$$m_2 = 4, \quad m_2' = -12, \quad m_3 = -18, \tag{3.15}$$

so that from the appendix,

$$f_1 = f_2 = v, \quad f_3 = 1 + 2v + 3y + y_5, \quad f_4 = y, \quad f_5 = 4v + 2,$$

$$f_6 = 1 + 2v^2 + 6y + 2y_5,$$

$$f_7 = 1 + 8v + 4v^2 + 6y, \quad f_8 = 6v, \quad f_9 = 24(v + v^2) + 18y,$$

$$f_{10} = 24(2v^2 + y), \tag{3.16}$$

where

$$v = v_2 - 2v_1, \quad y = 4y_1 - 2y_2 + y_3 + 6y_4. \tag{3.17}$$

From (3.15), (3.17) and the appendix, the values of v, y and y_5 for the LR, Rao's and Wald's statistics are given by

$$v^{\mathrm{LR}} = -\tfrac{1}{3}, \quad y^{\mathrm{LR}} = \tfrac{7}{36}, \quad y_5^{\mathrm{LR}} = -\tfrac{1}{2}, \quad v^{\mathrm{Rao}} = y^{\mathrm{Rao}} = 0, \quad y_5^{\mathrm{Rao}} = -\tfrac{1}{2},$$

$$v^{\mathrm{Wald}} = -1, \quad y^{\mathrm{Wald}} = 1, \quad y_5^{\mathrm{Wald}} = -\tfrac{1}{2}. \tag{3.18}$$

By (3.7), (3.9) and (3.16),

$$S(z) = 2z(1 + vz^2), \quad U(z) = -z + (1 - 6v)z^3 + v(2 - v)z^5. \tag{3.19}$$

For the LR, Rao's and Wald's statistics, using (3.18), these expressions reduce to

$$S^{\mathrm{LR}}(z) = 2z(1 - \tfrac{1}{3}z^2), \quad S^{\mathrm{Rao}}(z) = 2z, \quad S^{\mathrm{Wald}}(z) = 2z(1 - z^2),$$

$$U^{\mathrm{LR}}(z) = -z + 3z^3 - \tfrac{7}{9}z^5, \quad U^{\mathrm{Rao}}(z) = -z + z^3, \quad U^{\mathrm{Wald}}(z) = -z + 7z^3 - 3z^5.$$

Hence, by (3.8), with regard to second-order local maximinity, Rao's statistic will be superior to the LR statistic if $z^2 > 6$ and Wald's statistic if $z^2 > 2$. Also, comparing the expressions for $U(z)$ for the three statistics, it is seen that under the criterion of third-order local average power Rao's statistic will be better than the LR statistic if $z^2 > \frac{18}{7}$ and Wald's statistic if $z^2 > 2$. In particular, for $\alpha = 0.05$ (i.e., $z = 1.96$), Rao's statistic will dominate Wald's statistic under both the criteria and the LR statistic under the criterion of third-order local average power.

From (3.19), we also note that there does not exist a choice of v that maximizes $U(z)$ uniformly in z.

Remark. While Rao's statistic turns out to be quite attractive in the above example, the conclusion can be different with other estimating functions. For example, if one works with the usual likelihood then the LR statistic is Bartlett adjustable and always at least as good as Rao's statistic in terms of second-order local maximinity — see Mukerjee (1993).

4. Concluding remarks

The present work is the first attempt to understand the higher-order asymptotics for test statistics arising from quasi-likelihood and one of our primary objectives has been to demonstrate that the results can differ considerably from those based on the usual likelihood. With heavier notation and algebra, it should be possible to extend our results to the case where nuisance parameters are present but, in the absence of the relevant Bartlett identities, the resulting expressions would be very messy.

For discrete $X^{(n)}$, Edgeworth assumptions do not hold and hence interpretation of moment expansions as in this paper becomes problematic. However, a possible route sketched in Ghosh (1994, Chapter 9) may lead to similar results without Edgeworth assumptions.

Before concluding, we briefly indicate the prospects for a Bayesian version of our results on Bartlett adjustability. Considering, for example, the LR statistic T^{LR}, one can check that even in the iid case T^{LR} is distributed as a multiple of a non-central chi-square variate, up to the first order of approximation, in the posterior setup given $X^{(n)}$. Under mild conditions, this limiting distribution becomes central chi-square only if the quasi likelihood coincides with the usual likelihood. Hence, there is no point in investigating posterior Bartlett adjustability, given $X^{(n)}$, with reference to a quasi-likelihood. However, if we consider the posterior setup given $\hat{\theta}$, then the first-order posterior distribution of T^{LR} is again a central chi-square. We can also consider higher-order posterior expansions given the second and higher derivatives of the quasi-likelihood at $\hat{\theta}$ and investigate Bartlett adjustability. But no neat results are expected.

Acknowledgements

Thanks are due to professor O.E. Barndorff-Nielsen for suggesting the problem and to Professor A.K. Bera and the referees for very helpful comments. The work of RM was supported by a grant from the Centre for Management and Development Studies, Indian Institute of Management Calcutta.

Appendix

We show below certain expressions which have been used in the main text of the paper:

$$v_1^{\mathrm{LR}} = \tfrac{1}{2}I^{-3/2}, \quad v_2^{\mathrm{LR}} = \tfrac{1}{6}I^{-5/2}m_2,$$

$$y_1^{\mathrm{LR}} = \tfrac{3}{8}I^{-5/2}, \quad y_2^{\mathrm{LR}} = \tfrac{5}{12}I^{-7/2}m_2, \quad y_3^{\mathrm{LR}} = \tfrac{1}{9}I^{-9/2}m_2^2 + \tfrac{1}{24}I^{-7/2}m_3,$$

$$y_4^{\mathrm{LR}} = \tfrac{1}{6}I^{-5/2}, \quad y_5^{\mathrm{LR}} = -\tfrac{1}{2}I^{-3/2}\Delta,$$

$$v_1^{\mathrm{Rao}} = v_2^{\mathrm{Rao}} = y_1^{\mathrm{Rao}} = y_2^{\mathrm{Rao}} = y_3^{\mathrm{Rao}} = y_4^{\mathrm{Rao}} = 0, \quad y_5^{\mathrm{Rao}} = -\tfrac{1}{2}I^{-3/2}\Delta,$$

$$v_1^{\mathrm{Wald}} = I^{-3/2}, \quad v_2^{\mathrm{Wald}} = \tfrac{1}{2}I^{-5/2}(m_2 + I'),$$

$$y_1^{\mathrm{Wald}} = I^{-5/2}, \quad y_2^{\mathrm{Wald}} = I^{-7/2}(\tfrac{3}{2}m_2 + I'),$$

$$y_3^{\mathrm{Wald}} = I^{-7/2}(\tfrac{1}{6}m_3 + \tfrac{1}{4}I'') + I^{-9/2}\{\tfrac{1}{2}m_2^2 + \tfrac{1}{2}m_2 I' - \tfrac{1}{8}(I')^2\},$$

$$y_4^{\mathrm{Wald}} = \tfrac{1}{2}I^{-5/2}, \quad y_5^{\mathrm{Wald}} = -\tfrac{1}{2}I^{-3/2}\Delta,$$

$$f_1 = v_1 J + v_2 I, \quad f_2 = I^{-1/2}(\tfrac{1}{2}m_2 + I') - v_1 I(m_2 + I') + v_2 I^2,$$

$$f_3 = I^{-1/2}\Delta + v_1(J' - L - M) + v_2(I' - 2J) + y_1\{IM - 2(m_2 + I')J\}$$
$$+ y_2\{2IJ - (m_2 + I')I\} + 3y_3 I^2 + y_4\{2L - (m_3 - m_2')\}I + y_5 I,$$

$$f_4 = \tfrac{1}{2}I^{-1/2}(m_2' + I'' - \tfrac{1}{3}m_3) - v_1\{(m_2 + I')(\tfrac{1}{2}m_2 + I') + I(m_2' + \tfrac{1}{2}I'' - \tfrac{1}{2}m_3)\}$$
$$+ v_2 I(m_2 + 2I') + y_1 I(m_2 + I')^2 - y_2 I^2(m_2 + I') + y_3 I^3 - y_4 I^2(m_3 - m_2'),$$

$$f_5 = I^{-1}(I' - 2J) + 2v_1 I^{1/2}(J - m_2 - I') + 4v_2 I^{3/2},$$

$$f_6 = 2I^{-1/2}\{v_1 N + v_2 K + y_1(IM + 2J^2) + 3y_2 IJ + 3y_3 I^2 + 3y_4 IL + y_5 I\}$$
$$+ v_1^2(IM + J^2) + 2v_2^2 I^2 + 4v_1 v_2 IJ + I^{-1}\Delta,$$

$$f_7 = I^{-1}(\tfrac{1}{2}I'' + L + M - 2J') + v_1 I^{-1/2}\{2I'(3J - I') + m_2(5J - 2I')$$
$$+ I(2J' - 2L - 2M - I'' - 2m_2' + m_3)\} + 2v_2 I^{1/2}(4I' - 4J + m_2)$$
$$+ v_1^2 I\{IM + (m_2 + I')(m_2 + I' - 2J)\}$$
$$+ 4v_2^2 I^3 + 4v_1 v_2 I^2(J - m_2 - I') + 2y_1 I^{1/2}(m_2 + I')(m_2 + I' - 2J)$$
$$+ 2y_2 I^{3/2}(J - 2m_2 - 2I') + 6y_3 I^{5/2} + 2y_4 I^{3/2}(L - 2m_3 + 2m_2'),$$

$$f_8 = I^{-3/2}K + 6v_1 J + 6v_2 I,$$

$$f_9 = I^{-3/2}(K' - 3N) + 3v_1[2J' + N - 2L - 2M + I^{-1}\{2JI' - 4J^2 - K(m_2 + I')\}]$$
$$+ 6v_2(2I' + K - 4J) + 6v_1^2 I^{1/2}\{IM + J^2 - 2J(m_2 + I')\} + 24v_2^2 I^{5/2}$$
$$+ 12v_1v_2I^{3/2}(3J - m_2 - I') + 6y_1J(J - 2m_2 - 2I') + 6y_2I(2J - m_2 - I')$$
$$+ 18y_3I^2 + 6y_4I(2L - m_3 + m_2'),$$

$$f_{10} = I^{-2}H + 12v_1I^{-3/2}(JK + IN) + 24v_2I^{-1/2}K + 12v_1^2(IM + 3J^2) + 48v_2^2I^2$$
$$+ 96v_1v_2IJ + 24I^{-1/2}(y_1J^2 + y_2IJ + y_3I^2 + y_4IL).$$

References

Barndorff-Nielsen, O.E., 1995. Quasi profile and directed likelihoods from estimating functions. Ann. Inst. Statist. Math. 47, 461–464.

Bickel, P.J., Ghosh, J.K., 1990. A decomposition for the likelihood ratio statistic and the Bartlett correction – a Bayesian argument. Ann. Statist. 18, 1070–1090.

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

Chandra, T.K., Joshi, S.N., 1983. Comparison of the likelihood ratio, Rao's and Wald's tests and a conjecture of C.R. Rao. Sankhyā A 45, 226–246.

Chandra, T.K., Mukerjee, R., 1985. Comparison of the likelihood ratio, Wald's and Rao's tests. Sankhyā A 47, 271–284.

Ghosh, J.K., 1994. Higher Order Asymptotics. IMS and ASA, Hayward, CA.

McCullagh, P., 1991. Quasi likelihood and estimating equations. In: Hinkley, D.V., Reid, N., Snell, E.J. (Eds.), Statistical Theory and Modelling. Chapman and Hall, London, (chapter 11).

Mukerjee, R., 1992. Conditional likelihood and power: higher order asymptotics. Proc. Roy. Soc. London A 438, 433–446.

Mukerjee, R., 1993. Rao's score test: recent asymptotic results. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics. 11. North-Holland, Amsterdam, pp. 363–379.

Rao, C.R., 1973. Linear Statistical Inference and its Applications, 2nd Edition. Wiley, New York.

Taniguchi, M., 1991. Third order asymptotic properties of a class of test statistics under a local alternative. J. Mult. Anal. 37, 223–238.

Viraswami, K., Reid, N., 1996a. Higher order asymptotics under model misspecification. Preprint.

Viraswami, K., Reid, N., 1996b. Higher order results on the likelihood ratio statistic and its adjusted version under model misspecification. Preprint.