Adjusted *versus* Conditional Likelihood: Power Properties and Bartlett-type Adjustment

By J. K. GHOSH†

Purdue University, West Lafayette, USA, and Indian Statistical Institute, Calcutta, India

and

RAHUL MUKERJEE

Indian Institute of Management, Calcutta, India

SUMMARY

Under contiguous alternatives the adjusted likelihood ratio (ALR) test is seen to have the same power properties as the conditional likelihood ratio test up to the third order of comparison. In particular, an optimum property of the ALR test in terms of second-order local maximinity follows. It is also seen that the ALR statistic admits a Bartlett-type adjustment.

Keywords: AVERAGE POWER; BARTLETT ADJUSTMENT; LOCAL MAXIMINITY; PARAMETRIC ORTHOGONALITY

1. INTRODUCTION

The problem of adjusting the usual profile likelihood in an effective manner to handle nuisance parameters has received considerable attention in recent years; see section 2 of McCullagh and Tibshirani (1990) for a review. Cox and Reid (1987) pioneered the idea of conditional likelihood and discussed many interesting features of it. Another significant contribution to this area has recently been made by McCullagh and Tibshirani (1990) who introduced the notion of adjusted likelihood, derived interesting results on it and raised several open issues. In particular, they posed problems relating to

(a) a comparison between adjusted and conditional likelihood possibly via asymptotic considerations and
(b) the existence of a Bartlett-type adjustment for the adjusted likelihood ratio (ALR) statistic.

The present work attempts to settle some of these problems. It has also been noted that the desirable properties of the ALR test continue to hold even when one adjusts only the mean but not the variance of the score function.

2. POWER PROPERTIES

Let $\{X_i\}$, $i \geqslant 1$, be a sequence of independent and identically distributed random variables with common density $f(x; \theta, m)$ where θ is the parameter of interest and m is the nuisance parameter, both one dimensional. The parameter space is an open subset of $\mathbf{R}^2$. Interest lies in testing H_0: $\theta = \theta_0$ against $\theta \neq \theta_0$. Let n be the sample size and m_0^*

be the maximum likelihood estimator (MLE) of m given $\theta=\theta_0$. Write $D_1 = \partial/\partial\theta$, $D_2 = \partial/\partial m$ and for $i, j, i', j' = 0, 1, 2, \ldots$ define $K_{ij} \equiv K_{ij}(\theta, m) = E_{\theta, m}\{D_1^i D_2^j \log f(X; \theta, m)\}$, $A_{20} = -K_{20}(\theta, m)$, $A_{02} = -K_{02}(\theta, m)$,

$$K_{ij.i'j'} \equiv K_{ij.i'j'}(\theta, m) = E_{\theta, m}\{D_1^i D_2^j \log f(X; \theta, m) D_1^{i'} D_2^{j'} \log f(X; \theta, m)\},$$

$$H_{ij}(\theta, m) = n^{-1/2} \sum_{s=1}^{n} \{D_1^i D_2^j \log f(X_s; \theta, m) - K_{ij}(\theta, m)\},$$

$H_{ij}^* = H_{ij}(\theta_0, m_0^*)$, $K_{ij}^* = K_{ij}(\theta_0, m_0^*)$, $K_{ij.i'j'}^* = K_{ij.i'j'}(\theta_0, m_0^*)$, $a_{20}^* = -K_{20}^*$, $a_{02}^* = -K_{02}^*$, $L_{ij} = K_{ij}(\theta_0, m)$, $L_{ij.i'j'} = K_{ij.i'j'}(\theta_0, m)$, $a_{20} = -L_{20}$, $a_{02} = -L_{02}$.

$K_{ij.i'j'.i''j''}(\theta, m)$, $L_{ij.i'j'.i''j''}$ etc. are defined similarly. Among the K_{ij}, $K_{ij.i'j'}$ etc. only those which are used in what follows are assumed to exist; it is also assumed that they are smooth functions of θ and m. Since θ and m are one dimensional, following Cox and Reid (1987), we assume global parametric orthogonality, i.e. $K_{10.01}(\theta, m) \equiv 0$, identically in θ, m. Then under standard regularity conditions the per observation information matrix at (θ, m) is $\mathrm{diag}(A_{20}, A_{02})$ which is assumed to be positive definite for each (θ, m).

Following section 3 of McCullagh and Tibshirani (1990), the ALR statistic for H_0: $\theta=\theta_0$ is defined as $\lambda_{an} = 2\{l_{ap}(\tilde\theta) - l_{ap}(\theta_0)\}$, where

$$\left.\begin{aligned} l_{ap}(\theta) &= \int^{\theta} \bar U(t)\,\mathrm{d}t, \qquad \bar U(\theta) = \{U(\theta) - \mu(\theta)\}\, w(\theta), \qquad U(\theta) = D_1 l_p(\theta), \\ \mu(\theta) &= E_{\theta, \hat m_\theta} U(\theta), \qquad w(\theta) = [-E_{\theta, \hat m_\theta}\{D_1^2 l_p(\theta)\} + D_1 \mu(\theta)]/\mathrm{var}_{\theta, \hat m_\theta}\{U(\theta)\}, \\ l_p(\theta) &= l(\theta, \hat m_\theta), \qquad l(\theta, m) = \sum_{s=1}^{n} \log f(X_s; \theta, m), \end{aligned}\right\} \quad (2.1)$$

$\hat m_\theta$ is the MLE of m given θ and $\tilde\theta$ satisfies $l_{ap}(\tilde\theta) = \sup_\theta\{l_{ap}(\theta)\}$. We consider contiguous alternatives of the form $\theta_n = \theta_0 + n^{-1/2}\delta$. Since under parametric orthogonality $\mu(\theta) = -\xi_1(\theta, \hat m_\theta) + o(n^{-1/2})$, $w(\theta) = 1 - n^{-1}\xi_2(\theta, \hat m_\theta) + o(n^{-1})$, where

$$\xi_1(\theta, m) = \tfrac{1}{2} A_{02}^{-1} K_{12},$$

$$\xi_2(\theta, m) = (A_{20}A_{02})^{-1}(K_{22} + \tfrac{3}{2}K_{12.10} + K_{01.21} + 2K_{11.11} + 2K_{10.01.11} - \tfrac{1}{2}A_{02}^{-1}K_{12}K_{10.02})$$

(see McCullagh and Tibshirani (1990)), calculations similar to those in Mukerjee and Chandra (1991) show that $\tilde\theta = \hat\theta + \frac{1}{2}n^{-1}(a_{20}^* a_{02}^*)^{-1}K_{12}^* + o(n^{-1})$, and

$$\lambda_{an} = W_{an}^2 + o(n^{-1}), \quad (2.2a)$$

over a set with $P_{\theta_n, m}$-probability $1 + o(n^{-1})$ uniformly over compact subsets of δ, where $\hat\theta$ is the MLE of θ, and

$$W_{an} = (a_{20}^*)^{-1/2} H_{10}^* + n^{-1/2} Q_{1a} + n^{-1} Q_{2a}, \quad (2.2b)$$

$$Q_{1a} = \frac{1}{2}(a_{20}^*)^{-3/2} H_{10}^* H_{20}^* + \frac{1}{6}(a_{20}^*)^{-5/2} K_{30}^* H_{10}^{*2} + \frac{1}{2}(a_{20}^*)^{-1/2}(a_{02}^*)^{-1} K_{12}^*, \quad (2.2c)$$

$$Q_{2a} = \frac{3}{8}(a_{20}^*)^{-5/2} H_{10}^* H_{20}^{*2} + \frac{5}{12}(a_{20}^*)^{-7/2} K_{30}^* H_{10}^{*2} H_{20}^* + \frac{1}{6}(a_{20}^*)^{-5/2} H_{10}^{*2} H_{30}^*$$

$$+\left[\frac{1}{24}(a^*_{20})^{-7/2}\{K^*_{40}+3(a^*_{02})^{-1}K^{*2}_{21}\}+\frac{1}{9}(a^*_{20})^{-9/2}K^{*2}_{30}\right]H^{*3}_{10}$$

$$+\frac{1}{2}(a^*_{20})^{-3/2}(a^*_{02})^{-1}H^*_{10}H^{*2}_{11}+\frac{1}{2}(a^*_{20})^{-5/2}(a^*_{02})^{-1}K^*_{21}H^{*2}_{10}H^*_{11}$$

$$+(a^*_{20})^{-3/2}(a^*_{02})^{-1}\left[\frac{1}{4}K^*_{12}H^*_{20}+\left\{\frac{1}{6}(a^*_{20})^{-1}K^*_{12}K^*_{30}-\frac{1}{4}K^*_{22}-K^*_{10.01.11}\right.\right.$$

$$\left.\left.-\frac{1}{2}(K^*_{12.10}+K^*_{01.21})-K^*_{11.11}+\frac{1}{2}(a^*_{02})^{-1}K^*_{12}\left(\frac{1}{2}K^*_{12}+K^*_{10.02}\right)\right\}H^*_{10}\right]. \quad (2.2d)$$

By equations (2.2), the ALR test belongs to the broad family considered in Mukerjee (1992) and, up to $o(n^{-1/2})$, the expansion for W_{an} is identical with that for the 'square-root' version of the conditional likelihood ratio (CLR) statistic given there. Hence, by the results in Mukerjee (1992), when compared at the same size up to $o(n^{-1})$, the ALR and CLR tests will have, under contiguous alternatives, identical power functions up to $o(n^{-1/2})$ and identical average power functions up to $o(n^{-1})$ where averaging is done over values of θ equidistant from θ_0. Consequently, the ALR test, like the CLR test, will be optimal in terms of second-order local maximinity within the large class of tests considered in Mukerjee (1992). Also, for fixed m, the power function of the ALR test, like that of the CLR test, will be identical, up to $o(n^{-1/2})$, with the power function of the LR test with known nuisance parameter. Following Mukerjee (1992), we can work out conditions under which the ALR test is superior to the usual LR test with regard to third-order average power and satisfy ourselves that such conditions are satisfied in many examples of interest.

It is interesting to investigate the consequences of adjusting only the mean but not the variance of the score function. Then an ALR statistic can be defined as before with the change that the factor $w(\theta)$ will not occur in $\tilde{U}(\theta)$ (see equations (2.1)). Its square-root version will be as given by equation (2.2b) with the expression for Q_{ia} unchanged. Hence, following Mukerjee (1992), an ALR test, arising from such a simple adjustment of the profile likelihood, will continue to enjoy the desirable properties discussed above—see Ferguson *et al.* (1991). Thus, if interest lies in power properties as considered here, then the variance adjustment of McCullagh and Tibshirani (1990) is not essential.

3. BARTLETT-TYPE ADJUSTMENT

Starting from equations (2.2a)–(2.2d) we can make a further expansion about (θ_0, m) and proceed along the lines of Mukerjee and Chandra (1991) to show that the ALR statistic, like the CLR statistic, admits a Bartlett-type adjustment and that the Bartlett adjustment factor is given by $1+n^{-1}B_a(m^*_0)$, where

$$B_a(m)=a_{20}^{-3}\left(\frac{1}{36}L^2_{10.10.10}+\frac{7}{4}L^2_{10.20}+\frac{11}{6}L_{30}L_{10.20}+\frac{7}{18}L^2_{30}\right)-(a^2_{20}a_{02})^{-1}L_{21}\left(\frac{1}{4}L_{21}+L_{20.01}\right)$$

$$+a_{20}^{-2}\left(\frac{1}{4}L_{40}+L_{20.20}+L_{10.30}+L_{10.10.20}\right)-(a_{20}a^2_{02})^{-1}L_{21}\left(L_{01.02}+\frac{1}{2}L_{03}\right). \quad (3.1)$$

Comparing with Mukerjee and Chandra (1991), the Bartlett adjustment factors for the ALR and CLR statistics are identical if and only if $\{D_2K_{21}\}_{\theta=\theta_0} + \frac{1}{2}a_{02}^{-1}L_{03}L_{21} = 0$. Following Mukerjee (1992), even after a Bartlett-type adjustment a test based on the ALR statistic will enjoy the desirable properties discussed above. We can also check that if only the mean but not the variance of the score function is adjusted then the resulting version of the ALR statistic will admit a Bartlett-type adjustment. A consideration of the exponential regression model with the regression slope as the parameter of interest reveals that the Bartlett adjustment factor for this simpler version (and also for the original version) of the ALR statistic is not necessarily identical with that for the CLR statistic.

Proceeding as in Ghosh and Mukerjee (1992) and under the assumptions stated in their section 2, it can be seen that except for sample points with probability of the order $O(n^{-2})$ the ALR statistic has a posterior distribution, with reference to a prior density $\pi(\ )$ which is positive and thrice continuously differentiable, such that a posterior Bartlett adjustment exists. Furthermore, by equation (3.1), posterior probability regions based on a posterior Bartlett adjusted ALR statistic will have frequentist validity up to $o(n^{-1})$ if and only if $\pi(\ )$ satisfies

$$D_1\{A_{20}^{-1}D_1\,\pi(\theta,m) - A_{20}^{-2}K_{10,20}\,\pi(\theta,m)\} + D_2(A_{20}A_{02})^{-1}K_{21}\,\pi(\theta,m) = 0, \quad (3.2)$$

which is precisely the same as the corresponding condition, derived in Ghosh and Mukerjee (1992), for the CLR statistic. We refer to Ghosh and Mukerjee (1992) for a discussion on the availability of solutions to equation (3.2).

ACKNOWLEDGEMENTS

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

REFERENCES

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

Ferguson, H., Reid, N. and Cox, D. R. (1991) Estimating equations from modified profile likelihood. In *Estimating Functions* (ed. V. P. Godambe), pp. 279–294. Oxford: Clarendon.

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

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

Mukerjee, R. (1992) Conditional likelihood and power: higher order asymptotics. *Proc. R. Soc.* 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.