Higher Order Results on an adjusted version of the Likelihood Ratio Statistic under Model Misspecification

K. Viraswami
Department of Statistics and Actuarial Science
University of Waterloo, Waterloo
Ontario, Canada
and
N. Reid
Department of Statistics
University of Toronto, Toronto
Ontario, Canada

Summary

Higher order results under model misspecification are obtained for the likelihood ratio statistic and the adjusted likelihood ratio statistic. The density functions of the two statistics to $O(n^{-3/2})$ under the null hypothesis, are derived for the scalar case. In addition, an improved version of the adjusted likelihood ratio statistic is also provided.

Some key words: Higher order results; model misspecification; likelihood ratio statistic; adjusted likelihood ratio statistic

1. Introduction

Suppose that we have a set of independent identically distributed observations $y_1, y_2, \ldots, y_n$ of a random variable Y, from some unknown density $h(y)$ and we fit a parametric model $\{f(y;\theta); \theta \in \Theta\}$ to it. If the log-likelihood function for the data is given by

$$\ell(\theta) = \ell(\theta; y) = \sum_{i=1}^{n} \log f(y_i; \theta),$$

then $\theta(h)$ is defined to be the value $\theta \in \Theta$ which maximizes the Fraser information

$$F(\theta) = \int \log f(y;\theta) h(y) dy,$$

see Kent (1982). Equivalently $\theta(h)$ is the limit in probability of the maximum likelihood estimate $\hat{\theta}$ of θ, estimated under the model $\{f(y;\theta); \theta \in \Theta\}$ when the density is in fact $h(y)$. The likelihood ratio statistic for testing the null hypothesis $H_0 : \theta(h) = \theta_0$ versus $H_a : \theta(h) \neq \theta_0$: θ and θ_0 are $p \times 1$ vectors, is given by

$$W = W(\theta_0) = 2\{\ell(\hat{\theta}) - \ell(\theta_0)\}.$$

Under the assumption that $h(y)$ is a member of the parametric family $\{f(y;\theta); \theta \in \Theta\}$, the likelihood ratio statistic, the score statistic as well as the Wald statistic, have an asymptotic chi-square distribution with p degrees of freedom (Cox and Hinkley, 1974, Ch. 9). Higher order theory on W has been extensively studied by various authors under the above assumption. Improvement on the χ^2 approximation of W under the null hypothesis $\theta = \theta_0$, is carried out by looking at its expected value which is given by

$$E(W) = p\{1 + \frac{b_1}{n} + O(n^{-2})\},$$

where n is the sample size, p is the dimension of θ and b_1 is either constant or if not constant can be estimated consistently. The term $(1 + \frac{b_1}{n})^{-1}$ or any quantity asymptotically equivalent to it, is called a Bartlett factor. The Bartlett adjusted likelihood ratio statistic is defined by

$$W' = W(1 + \frac{b_1}{n})^{-1}, \tag{1}$$

and has expected value p to order $O(n^{-2})$. The densities of W and W' to $O(n^{-3/2})$ have been derived through different approaches by Lawley (1956), Hayakawa (1977) and Barndorff-Nielsen and Cox (1979, 1984). However Barndorff-Nielsen and Hall (1988) showed that W' in fact has a chi-square density to $O(n^{-2})$, even when the adjustment term b_1 involves parameter estimates determined by maximum likelihood.

If $h(y)$ does not belong to the parametric family $\{f(y;\theta); \theta \in \Theta\}$ but $\hat{\theta}$ is a consistent estimate of θ_0, Kent (1982) showed that the distributions of the likelihood ratio statistic, the score statistic and the Wald statistic are equivalent to that of $\sum \mu_i V_i$ where $V_1, V_2, \ldots, V_p$ are independent $\chi^2(1)$ variates and $\mu_i, i = 1, \ldots, p$ are the eigenvalues of the matrix $(H^{-1}J)$ with

$$J = \int u_1(\theta_0) u_1^T(\theta_0) h(y_1) dy_1 \text{ and } H = -\int \frac{\partial u_1(\theta_0)}{\partial \theta_0^T} h(y_1) dy_1,$$

where $u_1(\theta) = \frac{\partial \ell_1(\theta)}{\partial \theta}$ and $\ell_1(\theta)$ is the contribution of one observation to the log-likelihood function. In this case since $J \neq H$, the first Bartlett identity fails.

For convenience, Kent (1982) introduced robust versions of the score statistic and the Wald statistic, which have a chi-square distribution with p degrees of freedom under the null hypothesis. Stafford (1996) compared the score, Wald and likelihood ratio statistic to these robust versions under model correctness. Higher order results on the robust score test statistic and robust Wald test statistic, under model misspecification have been developed in Viraswami and Reid (1996), but no higher order results on the likelihood ratio statistic are available in this situation.

In this paper, we introduce an adjusted version of the likelihood ratio statistic which is robust, in the sense that under the null hypothesis it has an asymptotic chi-square distribution with 1 degree of freedom. In addition, we obtain higher order results under model misspecification on the likelihood ratio statistic and this adjusted version. Following Viraswami and Reid (1996), the type of model misspecification that we consider is that the actual density of the data does not belong to the fitted parametric model, but still satisfies $\theta(h) = \theta_0$. This hypothesis is of interest whenever $\theta(h)$ is directly interpretable under $h(y)$, for example as an expected value. In §2, we develop results under model misspecification for the scalar parameter case. The density function to $O(n^{-3/2})$ of our adjusted likelihood ratio statistic as well as that of the likelihood ratio statistic, are derived. In §3 and §4, we provide an improved version of the adjusted likelihood ratio statistic having a chi-square density to third order, and an example comparing the results obtained using the third order expansions of the distribution function of the adjusted likelihood ratio statistic and its improved version.

2. Third Order Densities In Scalar Parameter Case

We follow the notation used in section 2 of Viraswami and Reid (1996). Using the power notation of McCullagh (1987, Ch. 7), expected information quantities are denoted by

$$I'_{ijk}(\theta) = \int \{\frac{\partial l_1(\theta)}{\partial \theta}\}^i \{\frac{\partial^2 l_1(\theta)}{\partial \theta^2}\}^j \{\frac{\partial^3 l_1(\theta)}{\partial \theta^3}\}^k h(y)dy,$$

where $l_1(\theta) = \log f(y_1;\theta)$ has been defined in §1. When $\theta = \theta(h) = \theta_0$, we will denote $I'_{ijk}(\theta_0)$ by I'_{ijk}, thus in the notation of §1 and Kent (1982), $I'_{20} = J$ and $I'_{01} = -H$. The derivatives of the log likelihood function will be written in terms of $O_p(1)$ random variables and $O(1)$ quantities; for example

$$Z_{ijk} = \frac{1}{n^{1/2}} \sum_{m=1}^{n} [\{\frac{\partial l_m(\theta)}{\partial \theta}\}^i \{\frac{\partial^2 l_m(\theta)}{\partial \theta^2}\}^j \{\frac{\partial^3 l_m(\theta)}{\partial \theta^3}\}^k - I'_{ijk}]|_{\theta=\theta_0}$$

is an $O_p(1)$ random variable with zero mean.

The asymptotic expansions for the densities of the scalar likelihood ratio statistic W, and its adjusted version will be obtained. In the first case, we will obtain an Edgeworth expansion for the density of the signed square root of the likelihood ratio statistic,W_1, and then use the moment generating function technique to obtain the required result.

An expansion of W_1 in terms of $O_p(1)$ random variables and $O(1)$ quantities is obtained from the expansion of W, which is derived in McCullagh (1987, Ch. 7). Expressions for the first five cumulants of W_1 are given in appendix 1. We note that k_5 is of order $O(n^{-3/2})$ and the higher cumulants are descending in order by $n^{-1/2}$, thus to order $O(n^{-1})$, the cumulant generating function of W_1 is written using its first four cumulants only. Its moment generating function is then obtained by exponentiation and inverted to give the density function of W_1 to $O(n^{-3/2})$. If we let $V = (-I'_{20}/I'_{01})^{-1/2}W_1$, then $W = (-I'_{20}/I'_{01})V^2$. Using the moment generating function technique the density function of V^2 is obtained in the form:

$$\begin{aligned} f_{V^2}(v_1) &= g_{\chi^2(1)}(v_1) + \frac{1}{n}[e_1 g_{\chi^2(7)}(v_1) + e_2 g_{\chi^2(5)}(v_1) \\ &\quad + e_3 g_{\chi^2(3)}(v_1) + e_4 g_{\chi^2(1)}(v_1)] + O(n^{-3/2}), \end{aligned} \tag{2}$$

where

$$\begin{aligned} e_1 &= \frac{-5I'_{11}I'_{001}}{4(I'_{01})^3} + \frac{15(I'_{11})^2}{8(I'_{01})^2 I'_{20}} + \frac{5(I'_{001})^2 I'_{20}}{24(I'_{01})^4} \\ &\quad - \frac{5I'_{11}I'_{30}}{4I'_{01}(I'_{20})^2} + \frac{5I'_{001}I'_{30}}{12(I'_{01})^2 I'_{20}} + \frac{5(\rho'_3)^2}{24}, \\ e_2 &= \frac{3I'_{02}}{8(I'_{01})^2} + \frac{I'_{101}}{2(I'_{01})^2} + \frac{I'_{001}I'_{11}}{(I'_{01})^3} - \frac{21(I'_{11})^2}{8(I'_{01})^2 I'_{20}} \\ &\quad - \frac{(I'_{001})^2 I'_{20}}{24(I'_{01})^4} - \frac{I'_{0001}I'_{20}}{8(I'_{01})^3} - \frac{3I'_{21}}{4I'_{01}I'_{20}} + \frac{11I'_{11}I'_{30}}{4I'_{01}(I'_{20})^2} \\ &\quad - \frac{2I'_{001}I'_{30}}{3(I'_{01})^2 I'_{20}} - \frac{5(\rho'_3)^2}{8} + \frac{\rho'_4}{8}, \\ e_3 &= -\frac{I'_{02}}{4(I'_{01})^2} - \frac{I'_{101}}{2(I'_{01})^2} + \frac{I'_{001}I'_{11}}{4(I'_{01})^3} + \frac{5(I'_{11})^2}{8(I'_{01})^2 I'_{20}} \\ &\quad - \frac{(I'_{001})^2 I'_{20}}{6(I'_{01})^4} + \frac{I'_{0001}I'_{20}}{8(I'_{01})^3} + \frac{I'_{21}}{I'_{01}I'_{20}} - \frac{7I'_{11}I'_{30}}{4I'_{01}(I'_{20})^2} \\ &\quad + \frac{I'_{001}I'_{30}}{4(I'_{01})^2 I'_{20}} + \frac{5(\rho'_3)^2}{8} - \frac{\rho'_4}{4}, \\ e_4 &= -\frac{I'_{02}}{8(I'_{01})^2} + \frac{(I'_{11})^2}{8(I'_{01})^2 I'_{20}} - \frac{I'_{21}}{4I'_{01}I'_{20}} + \frac{I'_{11}I'_{30}}{4I'_{01}(I'_{20})^2} \\ &\quad - \frac{5(\rho'_3)^2}{24} + \frac{\rho'_4}{8}, \end{aligned}$$

and $g_{\chi^2(r)}(.)$ denotes a chi-square density with r degrees of freedom. The density function of the likelihood ratio statistic itself does not have a simple form.

If the first Bartlett identity holds, that is $I'_{20} = -I'_{01}$, but the second and third Bartlett identities do not hold, then the density function of W to $O(n^{-3/2})$ will be given by (2) with I'_{20} replaced by $-I'_{01}$. If none of the Bartlett identities hold then the adjusted likelihood ratio statistic

$$W'_a = -\frac{I'_{01}}{I'_{20}} W,$$

will now have its density to order $O(n^{-3/2})$, given by (2). A similar adjustment of the signed square root of the likelihood ratio statistic has been provided by Stafford (1996). We notice that to first order, W'_a has a chi-square distribution with 1 degree of freedom and thus might be easier to use. The distribution function of W'_a is given by

$$\begin{aligned} Pr(W'_a \leq v_1) &= P_1(v_1) + \frac{1}{n}\{e_1 P_7(v_1) + e_2 P_5(v_1) + e_3 P_3(v_1) \\ &\quad + e_4 P_1(v_1)\} + O(n^{-3/2}), \end{aligned} \tag{3}$$

where $P_i(v_1) = Pr(\chi^2(i) \leq v_1)$. Estimates of I'_{01} and I'_{20} can be used in (3): for example $\hat{I}_{01}(\hat{\theta}) = \frac{1}{n}\sum_{m=1}^{n} \frac{\partial^2 \ell_m(\theta)}{\partial \theta^2}|_{\theta=\theta_0}$ and $\hat{I}_{20}(\hat{\theta}) = \frac{1}{n}\sum_{m=1}^{n}(\frac{\partial \ell_m(\theta)}{\partial \theta})^2|_{\theta=\theta_0}$.

Since the Bartlett corrected likelihood ratio statistic has a chi-square distribution with 1 degree of freedom to order $O(n^{-2})$, under model correctness, we will determine whether a Bartlett corrected version of W'_a, under model failure, also satisfies this property. We define $W'_{ab} = W'_a(1+\frac{b'(\theta)}{n})^{-1}$, where $b'(\theta)$ is obtained from the expression of the expected value of W'_a. Its first four cumulants are obtained as

$$\begin{aligned} k_1 &= 1 + O(n^{-3/2}), \\ k_2 &= 2 + \frac{1}{n}\{\frac{-6I'_{21}}{I'_{01}I'_{20}} - \frac{8I'_{11}I'_{30}}{I'_{01}(I'_{20})^2} + \frac{3I'_{02}}{(I'_{01})^2} + \frac{4I'_{101}}{(I'_{01})^2} \\ &\quad + \frac{24(I'_{11})^2}{(I'_{01})^2 I'_{20}} + + \frac{14I'_{001}I'_{30}}{3(I'_{01})^2 I'_{20}} - \frac{22I'_{001}I'_{11}}{(I'_{01})^3} \\ &\quad - \frac{I'_{0001}I'_{20}}{(I'_{01})^3} + \frac{14(I'_{001})^2 I'_{20}}{3(I'_{01})^4} + \frac{I'_{40}}{(I'_{20})^2}\} + O(n^{-3/2}), \\ k_3 &= 8 + \frac{1}{n}\{\frac{-72I'_{21}}{I'_{01}I'_{20}} - \frac{156I'_{11}I'_{30}}{I'_{01}(I'_{20})^2} + \frac{36I'_{02}}{(I'_{01})^2} + \frac{48I'_{101}}{(I'_{01})^2} \\ &\quad + \frac{378(I'_{11})^2}{(I'_{01})^2 I'_{20}} + \frac{76I'_{001}I'_{30}}{(I'_{01})^2 I'_{20}} - \frac{324I'_{001}I'_{11}}{(I'_{01})^3} \end{aligned}$$

$$-\frac{12I'_{0001}I'_{20}}{(I'_{01})^3}+\frac{66(I'_{001})^2I'_{20}}{(I'_{01})^4}+\frac{10(I'_{30})^2}{(I'_{20})^3}+\frac{12I'_{40}}{(I'_{20})^2}\}+O(n^{-3/2}),$$

$$\begin{aligned}k_4 &= 48+\frac{1}{n}\{\frac{-864I'_{21}}{I'_{01}I'_{20}}-\frac{2592I'_{11}I'_{30}}{I'_{01}(I'_{20})^2}+\frac{432I'_{02}}{(I'_{01})^2}+\frac{576I'_{101}}{(I'_{01})^2}\\&+\frac{5616(I'_{11})^2}{(I'_{01})^2I'_{20}}+\frac{1152I'_{001}I'_{30}}{(I'_{01})^2I'_{20}}-\frac{4608I'_{001}I'_{11}}{(I'_{01})^3}\\&-\frac{144I'_{0001}I'_{20}}{(I'_{01})^3}+\frac{912(I'_{001})^2I'_{20}}{(I'_{01})^4}+\frac{240(I'_{30})^2}{(I'_{20})^3}+\frac{144I'_{40}}{(I'_{20})^2}\}+O(n^{-2}).\end{aligned}$$

Unfortunately, there does not seem to be a simple relationship between these expressions and those of the uncorrected likelihood ratio statistic W'_a. It is clear that only very specialized models would have all the $O(n^{-1})$ terms equal to zero. However, under model correctness the Bartlett identities can be used to simplify these expressions and verify that

$$\begin{aligned}k_1 &= 1+O(n^{-3/2}),\\k_2 &= 2+O(n^{-2}),\\k_3 &= 8+O(n^{-3/2}),\\k_4 &= 48+O(n^{-2}).\end{aligned}$$

In this case the above cumulants agree with the first four cumulants of a χ_1^2 random variable to the given order.

If we use the version of the adjusted likelihood ratio given by $\hat{W}'_a=-\frac{\hat{I}_{01}}{\hat{I}_{20}}W$, then the density function to $O(n^{-3/2})$ is again of the form (2) with different coefficients

$$\begin{aligned}f_{\hat{W}'_a}(w_a) &= g_{\chi^2(1)}(w_a)+\frac{1}{n}\{e_5g_{\chi^2(7)}(w_a)+e_6g_{\chi^2(5)}(w_a)\\&\quad+e_7g_{\chi^2(3)}(w_a)+e_8g_{\chi^2(1)}(w_a)\}+O(n^{-3/2}),\end{aligned}\tag{4}$$

where now

$$\begin{aligned}e_5 &= \frac{5(I'_{001})^2I'_{20}}{24(I'_{01})^4}-\frac{5I'_{001}I'_{30}}{6I'_{20}(I'_{01})^2}+\frac{5(\rho'_3)^2}{6},\\e_6 &= \frac{I'_{101}}{2(I'_{01})^2}-\frac{I'_{001}I'_{11}}{(I'_{01})^3}-\frac{(I'_{001})^2I'_{20}}{24(I'_{01})^4}\\&\quad-\frac{I'_{0001}I'_{20}}{8(I'_{01})^3}+\frac{4I'_{001}I'_{30}}{3(I'_{01})^2I'_{20}}-\frac{5(\rho'_3)^2}{4}-\frac{\rho'_4}{4},\\e_7 &= \frac{1}{4}-\frac{I'_{101}}{2(I'_{01})^2}+\frac{I'_{001}I'_{11}}{(I'_{01})^3}-\frac{(I'_{001})^2I'_{20}}{6(I'_{01})^4}\end{aligned}$$

$$
\begin{aligned}
& +\frac{I'_{0001}I'_{20}}{8(I'_{01})^3} - \frac{I'_{001}I'_{30}}{2(I'_{01})^2 I'_{20}} + \frac{(\rho'_3)^2}{2} + \frac{\rho'_4}{4}, \\
e_8 &= -\frac{1}{4} - \frac{(\rho'_3)^2}{12}.
\end{aligned}
$$

Using (4), the resulting distribution function of $\hat{W}'_a$ is

$$
\begin{aligned}
Pr(\hat{W}'_a \leq w_a) &= P_1(w_a) + \frac{1}{n}\{e_5 P_7(w_a) + e_6 P_5(w_a) + e_7 P_3(w_a) \\
&\quad + e_8 P_1(w_a)\} + O(n^{-3/2}), \qquad (5)
\end{aligned}
$$

where $P_i(w_a) = Pr(\chi^2(i) \leq w_a)$. The cumulants of the Bartlett corrected version of $\hat{W}'_a$, under model failure, do not agree with those of a χ^2_1 random variable even to $O(n^{-1})$. In practice I'_{ijk} can be replaced by any consistent estimates such as $\hat{I}_{ijk}(\theta_0)$ and $\hat{I}_{ijk}(\hat{\theta})$.

3. Improved Statistics

As can be seen from the above, when the first three Bartlett identities fail the density function of W'_a and $\hat{W}'_a$ are not Bartlett correctable. We can however use the polynomial approach of Cordeiro and Ferrari (1991) to obtain an improved version of $\hat{W}'_a$ having a chi-square density to third order. The density function of the adjusted likelihood ratio statistic (4) can be written in the form given in Harris (1985) with

$$
\begin{aligned}
A_1 &= 6 + \frac{12 I'_{101}}{(I'_{01})^2} - \frac{24 I'_{001} I'_{11}}{(I'_{01})^3} - \frac{3 I'_{0001} I'_{20}}{(I'_{01})^3} \\
&\quad - \frac{8 I'_{001} I'_{30}}{I'_{20}(I'_{01})^2} + \frac{9 (I'_{001})^2 I'_{20}}{(I'_{01})^4} + 12(\rho'_3)^2 - 6\rho_4, \\
A_2 &= \frac{12 I'_{101}}{(I'_{01})^2} - \frac{24 I'_{001} I'_{11}}{(I'_{01})^3} - \frac{3 I'_{0001} I'_{20}}{(I'_{01})^3} - \frac{28 I'_{001} I'_{30}}{I'_{20}(I'_{01})^2} \\
&\quad + \frac{14 (I'_{001})^2 I'_{20}}{(I'_{01})^4} + 30(\rho'_3)^2 - 6\rho'_4, \\
A_3 &= -\frac{20 I'_{001} I'_{30}}{I'_{20}(I'_{01})^2} + \frac{5 (I'_{001})^2 I'_{20}}{(I'_{01})^4} + 20(\rho'_3)^2.
\end{aligned}
$$

Thus, using equation 3 in Cordeiro and Ferrari (1991) an improved version of the adjusted likelihood ratio statistic which has a chi-square distribution to $O(n^{-3/2})$ is

$$
\hat{W}^*_a = \hat{W}'_a[1 - \frac{1}{n}\{c + b\hat{W}'_a + a(\hat{W}'_a)^2\}],
$$

where

$$
\begin{aligned}
a &= -\frac{I'_{001}I'_{30}}{9(I'_{01})^2 I'_{20}} + \frac{(I'_{001})^2 I'_{20}}{36(I'_{01})^4} + \frac{(\rho'_3)^2}{9}, \\
b &= \frac{I'_{101}}{3(I'_{01})^2} - \frac{2I'_{001}I'_{11}}{3(I'_{01})^3} - \frac{I'_{0001}I'_{20}}{12(I'_{01})^3} + \frac{8I'_{001}I'_{30}}{9(I'_{01})^2 I'_{20}} \\
&\quad -\frac{(I'_{001})^2 I'_{20}}{36(I'_{01})^4} - \frac{5(\rho'_3)^2}{6} - \frac{\rho'_4}{6}, \\
c &= \frac{1}{2} + \frac{(\rho'_3)^2}{6}.
\end{aligned}
$$

4. Example

Suppose that we have a sample $y_1, y_2, \ldots, y_n$ from a Weibull distribution with shape parameter $s = 2$ and scale parameter $c = 1$. If we assume that the data is from a gamma distribution with shape parameter $\alpha = \Gamma(1.5)$ and scale parameter β, then to test the hypothesis $\beta = 1$, the adjusted likelihood ratio statistic and its improved version are given by

$$
\hat{W}'_a = \frac{\sum_{i=1}^n \{2y_i - \Gamma(1.5)\}}{\sum_{i=1}^n \{y_i - \Gamma(1.5)\}^2} \{2y_i - 2\Gamma(1.5) - 2\Gamma(1.5)\log(\bar{y}) + 2\Gamma(1.5)\log(\Gamma(1.5))\},
$$

$$
\hat{W}^*_{a1} = \hat{W}'_a \{1 - \frac{1}{n}(\hat{c}_1 + \hat{b}_1 \hat{W}'_a + \hat{a}_1 (\hat{W}'_a)^2)\},
$$

$$
\hat{W}^*_{a2} = \hat{W}'_a \{1 - \frac{1}{n}(\hat{c}_2 + \hat{b}_2 \hat{W}'_a + \hat{a}_2 (\hat{W}'_a)^2)\},
$$

where $\hat{a}_1, \hat{b}_1$ and $\hat{c}_1$ are obtained using expected information quantities, and $\hat{a}_2, \hat{b}_2$ and $\hat{c}_2$ are obtained using estimated information quantities $\hat{I}_{ijk}(\hat{\theta})$. For example $\hat{I}_{20}(\hat{\theta}) = \frac{1}{n}\sum(y_i - \frac{\bar{y}}{\Gamma(1.5)})^2$, $\hat{I}_{30} = \frac{1}{n}\sum(y_i - \frac{\bar{y}}{\Gamma(1.5)})^3$ and so on. We simulated data of sample size 20 from a Weibull $(2, 1)$ density. For each simulated sample, we computed $\hat{W}'_a, \hat{W}^*_{a1}, \hat{W}^*_{a2}$ and the estimated information quantities that are required. The exact and approximate values of $pr(\hat{W}'_a \leq w)$, $pr(\hat{W}^*_{a1} \leq w_1)$ and $pr(\hat{W}^*_{a2} \leq w_2)$ for $n = 20$, are provided in tables 1a, 1b and 1c. The exact values were obtained using 50,000 simulations.

When the 3rd and 4th likelihood moments are estimated from the data, the approximations are not better than the first order approximations. The approximations using exact information quantities are quite satisfactory. Using the improved versions of the adjusted likelihood ratio statistic, we obtain better approximations using exact moments.

5. Conclusions

Numerically, an adjusted version of the signed root of the likelihood ratio statistic (Stafford (1996)) is better than an adjusted version of the likelihood ratio statistic, under model failure. However, an explicit formulation of the adjusted version of the likelihood ratio statistic may be useful in comparing other approaches, and also as a reference for results in the nuisance parameter case.

Unfortunately it is not easy to extend our results to p-dimensional case because a simple multiplicative adjustment to the likelihood ratio statistic does not suffice. It is possible in principle to formulate an adjusted statistic by working with $W_r I^{r,s} W_s$, where $I^{r,s}$ is the (r, s)th element of the inverse of the covariance of the score function, W_r and W_s are as defined by equation (7.15) in McCullagh (1989). But as a referee has pointed out, it is not clear that this would give any substantial improvement beyond the use of the robust score statistic.

Acknowledgement

The authors thank C. Genest and a referee for correcting some earlier work in the p-dimensional case. The first author also thanks the Natural Sciences and Engineering Research Council of Canada for financial support.

Table 1a: Exact, first order and third order approximations for $pr(\hat{W}'_a \leq w)$.

w	0.0028	0.4076	1.0319	2.3211	2.7603	3.0128	3.4553	4.4430	5.2408
1st	0.0426	0.4768	0.6903	0.8724	0.9034	0.9174	0.9369	0.9650	0.9779
3rd(1)	0.0426	0.4768	0.6903	0.8728	0.9040	0.9182	0.9381	0.9669	0.9805
3rd(2)	0.0426	0.4768	0.6904	0.8974	0.9038	0.9177	0.9684	0.9762	0.9793
Exact	0.0404	0.4646	0.6811	0.8800	0.9140	0.9303	0.9490	0.9775	0.9878

Table 1b: Exact and third order approximations for $pr(\hat{W}^*_{a1} \leq w_1)$.

w_1	0.0028	0.4055	1.0625	2.5510	3.0971	3.4197	3.9998	5.3615	6.5233
Third	0.0420	0.4758	0.6973	0.8898	0.9216	0.9356	0.9545	0.9794	0.9894
Exact	0.0411	0.4723	0.6898	0.8918	0.9236	0.9412	0.9594	0.9844	0.9914

Table 1c: Exact and third order approximations for $pr(\hat{W}^*_{a2} \leq w_2)$.

w_2	0.0028	0.4095	1.0567	3.0392	3.3047	4.1076	4.9057	5.5607	5.9449
Third	0.0421	0.4778	0.6960	0.9187	0.9309	0.9573	0.9732	0.9816	0.9852
Exact	0.0410	0.4662	0.6807	0.8964	0.9072	0.9364	0.9580	0.9688	0.9747

1st denotes first order approximations.
3rd(1) denotes third order approximations using expected information quantities.
3rd(2) denotes third order approximations using estimated information quantities.

Appendix 1

The first five cumulants of W_1 are:

$$\begin{aligned}
k_1 &= \frac{1}{2\sqrt{n}}[\frac{I'_{11}}{(I''_{01})^{3/2}} + \frac{I'_{001}I'_{20}}{3(I''_{01})^{5/2}}] + O(n^{-3/2}). \\
k_2 &= \frac{I'_{20}}{I''_{01}} + \frac{1}{n}[\frac{I'_{21}}{(I''_{01})^2} + \frac{I'_{001}I'_{30} + 3(I'_{20})^2 I'_{02} + 3(I'_{20})^2 I'_{101}}{3(I''_{01})^3} + \frac{7(I'_{11})^2}{4(I''_{01})^3} \\
&\quad + \frac{I'_{0001}(I'_{20})^2}{4(I''_{01})^4} + \frac{17 I'_{001}I'_{20}I'_{11}}{6(I''_{01})^4} + \frac{13(I'_{001})^2(I'_{20})^2}{18(I''_{01})^5}] + O(n^{-3/2}). \\
k_3 &= \frac{1}{\sqrt{n}}[\frac{I'_{30}}{(I''_{01})^{3/2}} + \frac{3I'_{20}I'_{11}}{(I''_{01})^{5/2}} + \frac{I'_{001}(I'_{20})^2}{(I''_{01})^{7/2}}] + O(n^{-3/2}). \\
k_4 &= \frac{1}{n}[\frac{I'_{40}}{(I''_{01})^2} + \frac{6I'_{20}I'_{21} + 6I'_{11}I'_{30}}{(I''_{01})^3} + \frac{3(I'_{20})^2 I'_{02} + 18 I'_{20}(I'_{11})^2 + 4(I'_{20})^2 I'_{101}}{(I''_{01})^4} \\
&\quad + \frac{4I'_{001}I'_{30}I'_{20}}{(I''_{01})^4} + \frac{18 I'_{001}(I'_{20})^2 I'_{11} + I'_{0001}(I'_{20})^3}{(I''_{01})^5} + \\
&\quad \frac{4(I'_{001})^2(I'_{20})^3}{(I''_{01})^6}] + O(n^{-2}). \\
k_5 &= O(n^{-3/2}).
\end{aligned}$$

References

Barndorff-Nielsen, O. E. & Cox, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. B*, **41**, 279-312.

Barndorff-Nielsen, O. E. & Cox, D. R. (1984a). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. B*, **46**, 483-495.

Barndorff-Nielsen, O. E. & Cox, D. R. (1984b). The effect of sampling rules on likelihood statistics. *Int. Statist. Rev.*, **52**, 309-326.

Barndorff-Nielsen, O. E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*, **75**, 374-378.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. London Ser. A*, **160**, 268-282.

Cordeiro, G. M. & Ferrari, S. L. P. (1991). A modified score test statistic having chi-squared distribution to order n^{-1}. *Biometrika*, **78**, 3, 573-582.

Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics.* Chapman and Hall, London.

Harris, P. (1985). An asymptotic expansion for the null distribution of the efficient score statistic. *Biometrika*, **72**, 653-660.

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

Kent, T.J. (1982). Robust properties of likelihood ratio tests. *Biometrika*, **69**, 19-27.

Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, **43**, 295-303.

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

Stafford, J. E. (1996). A robust adjustment of the profile likelihood. *Ann. Statist.*, **24**, No. 1, 336-352.

Viraswami, K. (1995). Higher order asymptotics and robustness. *Ph.D Thesis, University of Toronto.*

Viraswami, K. and Reid, N. (1996). Higher order asymptotics under model misspecification. *Can. J. Statist.*, **24**, No. 2, 263-278.