A Historical Sketch and Some New Results on the Improved Log Likelihood Ratio Statistic

J. L. JENSEN

University of Aarhus

ABSTRACT. In the first part we discuss the known results concerning the reduction of the error of the chi-squared approximation when using the Bartlett adjusted log likelihood ratio statistic. In the second part we state some new results for the case where the first four log likelihood derivatives have both continuous and lattice variables. The latter results are of interest for example in connection with censored life times and in logistic regression. Finally, in the last section we discuss how to perform the algebraic manipulations using REDUCE.

Key words: asymptotic expansion, Bartlett adjustment, logistic regression, mixed data

1. Introduction

Let \( l_n(\theta), \theta = (\theta^{(1)}, \theta^{(2)}) = (\theta_1, \ldots, \theta_{p_1}, \theta_{p_1+1}, \ldots, \theta_p) \), be the log likelihood function based on \( n \) observations. The log likelihood ratio test statistic \( W_n \) for testing \( \theta^{(2)} = \theta_0^{(2)} \) is

\[
W_n = 2 \max_{\theta} l_n(\theta) - \max_{\theta^{(1)}} l_n(\theta^{(1)}, \theta_0^{(2)}),
\]

and under a number of regularity conditions \( W_n \) is asymptotically \( \chi^2_{p-p_1} \), where \( \chi^2 \) denotes a chi-squared distribution on \( f \) degrees of freedom. A natural choice for improving the asymptotic chi-squared distribution is to consider the statistic

\[
\tilde{W}_n = (p - p_1) \frac{W_n}{EW_n},
\]

with \( EW_n \) evaluated at \( \theta = (\tilde{\theta}^{(1)}, \tilde{\theta}_0^{(2)}) \), where \( \tilde{\theta}^{(1)} \) is the estimate when \( \theta^{(2)} = \theta_0^{(2)} \). In general, however, it is not possible to find \( EW_n \) explicitly, and instead an expansion is used,

\[
EW_n = (p - p_1) \left\{ 1 + \frac{B(\tilde{\theta})}{n} + O(n^{-2}) \right\},
\]

and then one uses the Bartlett adjusted statistic \( WB_n \),

\[
WB_n = \frac{W_n}{1 + B(\tilde{\theta}^{(1)}, \tilde{\theta}_0^{(2)})/n}.
\]

This idea was first exploited in Bartlett (1937) in his classical test for homogeneity of variances, the argument given there being simply “to obtain more precisely the value of the \( \chi^2 \) approximation”.

A detailed study of likelihood ratio statistics in the multivariate normal distribution is made by Box (1949). He derives formally an infinite series for the distribution of \( W_n \) in terms of chi-squared distributions, and with the terms decreasing in powers of \( 1/n \). Also he notes that the adjusted statistic \( WB_n \) has cumulants agreeing with those of a \( \chi^2_{p-p_1} \), except for a term of order \( n^{-2} \). For the class of tests discussed in Box (1949) the cumulant transform of \( W_n \) is actually known and it is therefore possible to use \( \tilde{W}_n \) in (1.2). This is discussed in Jensen (1991) together with other approximations.

One of the basic papers in this area is the one of Lawley (1956). He quite generally derives a formula for \( B \) in (1.3) in terms of the mean values of the first four derivatives of \( l_n(\theta) \) and
where 

There usually for transformed likelihood directly, then (1978) lattice discussion is (1987). some chi-squared where it see that Hayakawa formula the stochastic derivatives with those are caused H(O) and McCullagh & Cox (1984). Also included in these recent ideas is conditioning on an exact or approximate ancillary statistic, see in particular Barndorff-Nielsen & Cox (1984).

A general expansion of the distribution function of $W_n$ in the i.i.d. case is given in Hayakawa (1977). He obtains, by formal manipulations,

$$P(W_n \leq w) = \chi^2_{p-p_1}(w) + \frac{1}{24n} \left\{ A_2 \chi^2_{p-p_1+4}(w) - (2A_2 - A_1) \chi^2_{p-p_1+2}(w) 
+ (A_2 - A_1) \chi^2_{p-p_1}(w) \right\} + O\left(\frac{1}{n}\right),$$

where $A_1$ and $A_2$ are parameter functions, and $\chi^2_d(w)$ is the distribution function of a chi-squared distribution with degrees of freedom equal to $d$. If $A_2 = 0$ one obtains from (1.5) that

$$P(WB_n \leq w) = \chi^2_{p-p_1}(w) + O\left(\frac{1}{n}\right), \quad \text{with} \quad WB_n = \frac{W_n}{1 + A_1/[12n(p-p_1)]}. \quad (1.6)$$

It has caused some discussion in the literature whether $A_2$ is in general zero, see e.g. Cordeiro (1987). This however follows on using the left hand side of (1.5) for calculating $EW_n$ and then comparing with the cumulant calculation of Lawley (1956). The expression (1.6), usually with $O(n^{-1})$ replaced by $O(n^{-3/2})$ or $O(n^{-2})$, has therefore become the main argument for using $WB_n$ instead of $W_n$.

As mentioned Hayakawa's calculations are only formal calculations. In the sections 1.1–1.3 we discuss the known results dealing with the validity of (1.6). We divide the discussion into the case of variables having a density w.r.t. Lebesgue measure, the case of lattice variables, and finally the mixed case where some of the variables are continuous and some are discrete. The last case is also the subject of section 2 where some general results are proved.

1.1. The case of variables satisfying a Cramèr condition

There are two very basic papers for the validity of (1.6), namely that of Bhattacharya & Ghosh (1978) and that of Chandra & Ghosh (1979). Bhattacharya & Ghosh do not deal with (1.6) directly, but consider Edgeworth expansions for certain estimators including maximum likelihood estimators. This is based on a transformation result for a statistic of the form

$$V_n = \sqrt{n}H(\bar{Z}), \quad \bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i \quad (1.7)$$

where $H(0) = 0$, $EZ = 0$ and $\partial H/\partial z(0)$ is non-null. If $Z_i$ satisfies a Cramèr condition and $H$ is sufficiently smooth, the Edgeworth expansion for the distribution of $\sqrt{n}\bar{Z}$ can be transformed to an Edgeworth expansion for the distribution of $V_n$. This result can be used for example for $\bar{\theta}$ by writing $\sqrt{n}(\bar{\theta} - \theta_0) = \sqrt{n}H(\bar{Z}) + \text{remainder}$, where the coordinates of $Z_i$ are $Z_i^x = D^x_i(\theta_0) - ED^x_i(\theta_0)$ for $1 \leq |v| \leq s$ with $v = (v_1, \ldots, v_p) \in \mathbb{Z}^p$, $|v| = v_1 + \ldots + v_p$, and $D^x = \partial^{|v|}\partial\theta_1 \ldots \partial\theta_p$. The statistic $\bar{Z}$ need not be exactly a standardized sum as
in (1.7), it suffices that \( \sqrt{n} Z \) has a valid Edgeworth expansion, see the general formulation in Skovgaard (1981).

Chandra & Ghosh (1979) extend these ideas by considering statistics of the form

\[
Q_n = nH(\bar{U}), \quad \bar{U} = \frac{1}{n} \sum_{i=1}^{n} U_i, \quad \frac{\partial H}{\partial z^*}(0) = 0, \quad LV(U_1)L = L, \tag{1.8}
\]

where \( L = \frac{\partial^2 H}{\partial z \partial z^*}(0) \) and \( V(U_1) \) is the variance of \( U_1 \). It is then proved that the log likelihood ratio statistic \( W_n \) may be written as \( W_n = nH(\bar{U}) + \text{remainder} \), where \( \bar{U} = (\hat{\theta} - \theta_0, Z) \) with \( Z \) as above. For completeness we extract the following theorem from Chandra & Ghosh (1979).

Let \( X_1, \ldots, X_n \) be i.i.d. variables on \( \mathbb{R}^m \) with distribution \( P_0 \), and let \( h(x; \theta) \), \( \theta \in \Theta \) an open set in \( \mathbb{R}^p \), be the logarithm of the density w.r.t. a dominating measure \( \mu \). The following assumptions are used:

(A1) There is an open set \( \mathcal{W} \subseteq \mathbb{R}^m \) such that \( P_0(\mathcal{W}) = 1 \) for all \( \theta \in \Theta \) and the \( v \)-th derivative \( D^v h(x; \theta) \) w.r.t. \( \theta \) exists on \( \mathcal{W} \times \Theta \) for all \( 1 \leq |v| \leq 4 \).

(A2) For all \( 1 \leq |v| \leq 4 \) we have \( E_\theta |D^v h(X_1; \theta)|^5 < \infty \), and there exists \( \epsilon = \epsilon(\theta) > 0 \) such that \( E_\theta \{ \max_{|\theta| < \epsilon} |D^v h(X_1; \theta)| \}^4 < \infty \) for \( |v| = 5 \).

(A3) The information matrix \( I(\theta) = (-E_\theta D_\theta D_\theta h(X_1; \theta)) \) is nonsingular, where \( D_\theta = \partial / \partial \theta \).

(A4) The functions \( I(\theta), E_\theta \{ (D^v h(X_1; \theta)(D^v h(X_1; \theta)) \}, 1 \leq |v|, |v'| \leq 4 \), are continuous on \( \Theta \).

(A5) \( P_\theta \) has an absolutely continuous component w.r.t. Lebesgue measure whose density is strictly positive on \( \mathcal{W} \). Also, for each \( v \), \( 1 \leq |v| \leq 4 \), \( D^v h(x; \theta) \) is continuously differentiable in \( x \) on \( \mathcal{W} \).

In section 2.2 in remark 2.6 we mention an alternative to (A5).

**Theorem 1.1**

Under the assumptions (A1) to (A5) there exists a constant \( A \) such that

\[
\sup_w \left[ P(W_n \leq w) - \left[ \chi^2_{p - p_1}(w) \left( 1 - \frac{A}{n} \right) + \frac{A}{n} \chi^2_{p - p_1 + 2}(w) \right] \right] = O(n^{-1}), \tag{1.9}
\]

which implies

\[
\sup_w \left| P(WB_n \leq w) - \chi^2_{p - p_1}(w) \right| = O(n^{-1}),
\]

where \( WB_n = W_n / \left[ 1 + 2A \left[ n(p - p_1) \right] \right] \).

**Remark 1.2.** It appears from Chandra & Ghosh (1979) that in general one gets an expansion in powers of \( n^{-1} \). Under the right moment assumptions the error term in theorem 1.1 can therefore be replaced by \( O(n^{-2}) \). This is also true for the mixed case below, but since the main point in this article is to discuss the validity of (1.9), I have not included the \( O(n^{-2}) \) error term in the discussion. Concerning what happens when unknown parameters in the coefficient \( A \) in (1.9) are replaced by estimated values see remark 2.10.

The above theorem is for the i.i.d. case, but the basic principle behind the result is an Edgeworth expansion for the first four derivatives of the log likelihood function and a subsequent transformation to \( W_n \). Thus for example Taniguchi (1988) has considered Gaussian ARMA processes and obtained an expansion for \( W_n \) for testing a simple hypothesis about the spectral density. Quite generally Edgeworth expansions can be obtained for sums of Markov dependent variables, see Jensen (1989a), but usually the coefficients of the expansions are hard to find.
For a signed square root of $W_n$ results comparable to those above have recently been established by Bickel & Ghosh (1990). Let $W_{nj}$ be the log likelihood ratio statistic for testing $	heta_j = 0$ under the model with $	heta_{j+1} = \ldots = \theta_p = 0$, and let $W_{nj} = V_{nj}^2$ with the sign determined by the estimate of $\theta_j$ under the same model. Then Bickel & Ghosh (1990) show that the $p$-variate statistic $(V_{n1}, \ldots, V_{np})$ can be bias and variance adjusted such that the error of the normal approximation becomes $O(n^{-3/2})$. Although the results are derived through a Bayesian setting they are essentially products of the results in Bhattacharya & Ghosh (1978).

The results for $(V_{n1}, \ldots, V_{np})$ in Barndorff-Nielsen (1986) are more restrictive in that they partly rely on a saddlepoint approximation. Barndorff-Nielsen (1986) also suggested an alternative statistic, named $r^*$ by him, with the same distributional properties as the bias and variance adjusted version of $V_{nj}$. For a direct derivation of $r^*$ from a saddlepoint approximation see Jensen (1992). The discussion in Barndorff-Nielsen (1986) as well as Barndorff-Nielsen & Cox (1984) is conditionally on an approximate ancillary statistic. I have chosen not to enter this discussion here partly because of the problem of finding an approximate ancillary statistic, and partly because the log likelihood ratio statistic $W_n$ is independent of an approximate ancillary statistic up to terms of order $O(n^{-1})$.

1.2. The lattice case

Let $X_i \in \{1, \ldots, p + 1\}$ be multinomial distributed with density

$$
\frac{dP_\theta}{d\mu}(x) = \exp \left\{ \theta_1 z_1 + \ldots + \theta_p z_p - \ln \left( 1 + \sum_{j=1}^p \exp(\theta_j) \right) \right\},
$$

where $z_j = 1_{ij}(x)$. The minimal sufficient statistic becomes $Z = (Z_1, \ldots, Z_p)$ with $Z_j = \sum Z_{ij}$, $Z_{ij} = 1_{ij}(X_i)$. This statistic has a lattice distribution and the method and results outlined in section 1.2 are not applicable. For the $\chi^2$ goodness of fit statistic for testing $\theta = \theta_0$,

$$
\chi^2 = \sum_{j=1}^{p+1} \frac{(Z_j - np_{jo})^2}{np_{jo}}
$$

with $Z_{p+1} = n - \sum_{j=1}^p Z_j$ and $p_{jo} = \exp(\theta_j)/(1 + \sum \exp(\theta_j))$, $p_{p+1,0} = (1 + \sum \exp(\theta_j))^{-1}$, Yarnold (1972) showed that the asymptotic chi-squared approximation is valid up to terms of order $O(n^{-p(p+1)})$ only. This is because the lattice structure of $Z$ implies that the distribution of the quadratic statistic $X^2$ has jumps larger than $O(n^{-1})$. It is tempting to think that the log likelihood ratio statistic will behave in a similar way, and thus we cannot obtain a result like (1.6) in the lattice case.

It has never been proved nor disproved that (1.6) holds in the lattice case, but Frydenberg & Jensen (1989) showed by extensive numerical calculations for $p$ small that (1.6) does not seem to hold. Their results point to the same conclusions as for the $X^2$-statistic considered by Yarnold (1972), and in general the Bartlett adjustment can both improve the chi-squared approximations as well as make it worse at a fixed significance level.

References to previous works for the lattice case can be found in Frydenberg & Jensen (1989).

As pointed out above the findings in Frydenberg & Jensen (1989) are for $p$ small. Contrary to this Götze (1989) has proved that (1.6) holds also in the lattice case if only the dimension $p$ is sufficiently large. In Götze’s result $p$ has to be of order 150, and it is an unsolved problem for which dimension $p$ that (1.6) changes from not being valid to being valid in the lattice case.
1.3. The mixed case

If we consider the exponential family where the minimal sufficient statistic has some coordinates with a continuous distribution and other coordinates with a lattice distribution the method of section 1.1 is not applicable. An example is provided by censored exponential life times with density

\[ \frac{dP_\theta}{d\mu}(x) = \exp \{-\theta x + \delta \ln \theta\}, \]

where \( 0 < x \leq T, \ T \) being the time of censoring, and \( \delta = 1_{(T)}(x) \) is the indicator function for censoring. For \( n \) observations the minimal sufficient statistic becomes \((\Sigma^i X_i, \Sigma^i \delta_i)\), with the second coordinate having a lattice distribution. This particular example was considered in Jensen (1987), and by first expanding the conditional distribution of \( \Sigma X_i \) given \( \Sigma \delta \), the result (1.6) was established for testing \( \theta = \theta_0 \).

The intuitive steps in establishing (1.6) in the mixed case are, that the conditional distribution of \( W_a \) given the lattice variables can be expanded if the signed square root of \( W_a \) depends functionally on the continuous part, and that the mean value of a smooth functional can be evaluated by replacing the true distribution by the formal Edgeworth expansion. This latter kind of results can be found in Götze & Hipp (1978). Thus a major step in this approach, parallel to the method of section 1.2, is a transformation result, this time for a statistic of the form \( \sqrt{n} H(X, Y) \), where the \( Y \)-part has a lattice distribution. Such a transformation result is established in Jensen (1989b), and we use this in section 2 to establish general results for estimators and test statistics in the mixed case.

2. Edgeworth expansions in the mixed case

2.1. Edgeworth expansions for estimators

We first extend a theorem of Bhattacharya & Ghosh (1978) on Edgeworth expansions for certain estimators, including the maximum likelihood estimator, to cases with both discrete and continuous variables.

Let \( X_1, X_2, \ldots \) be an i.i.d. sequence on the measure space \((\mathcal{X}, \mathcal{A})\) with common distribution \( P_\theta \), where \( \theta \) belongs to an open subset \( \Theta \) of \( \mathbb{R}^p \). For each \( \theta \) let \( h(x; \theta) \) be a real measurable function on \( \mathcal{X} \), which will play the role of a contrast function, see (2.2) below. For non-negative integral vectors \( v = (v_1, \ldots, v_p) \) write \( |v| = v_1 + \ldots + v_p \), \( v! = v_1! \ldots v_p! \) and \( D^r = \partial^{v_1} \partial^{v_2} \ldots \partial^{v_p} \). When \( |v| = 1 \) we write \( D_i \) instead of \( D^r \) when the derivative is w.r.t. \( \theta_i \). We shall use the following assumptions with \( s \geq 3 \) a fixed integer and \( \theta_0 \) the true parameter value.

(A1) For each \( v, 1 \leq |v| \leq s + 1 \), \( h(x; \theta) \) has a \( v \)-th derivative \( D^v h(x; \theta) \) with respect to \( \theta \) on \( \mathcal{X} \times \Theta \).

(A2) For each \( v, 1 \leq |v| \leq s \), \( E_{\theta_0} |D^v h(X_1; \theta_0)|^{s+1} < \infty \) and there exists \( a_i > 0 \) such that for each \( v, |v| = s + 1 \),

\[ E_{\theta_0} \left( \sup_{|0 - \theta_0| < a_i} |D^v h(X_1; \theta)|^s \right) < \infty. \]

(A3) \( E_{\theta_0} D_i h(X_1, \theta_0) = 0 \) for \( i = 1, \ldots, p \), and the \( p \times p \) matrices

\[ I(\theta_0) = \{-E_{\theta_0}[D D h(X_1; \theta_0)]\} \]

\[ D(\theta_0) = \{E_{\theta_0}[(D h(X_1; \theta_0))(D h(X_1; \theta_0))]\} \]

are non-singular.
These assumptions are the same as in Bhattacharya & Ghosh (1978) and basic to the problem. We shall then supplement them to handle the case of both discrete and continuous variables.

We define \( Z_i^{(1)} = D'h(X_i; \theta_0) \) and let \( Z_i = (Z_i^{(1)})_{1 \leq |i| \leq s} \) be the vector with coordinates indexed by \( i \)'s. The dimension of \( Z_i \) is \( k = \sum_{r=1}^{s} \left( \begin{array}{c} p + r - 1 \\ r \end{array} \right) \), and we let the numbering be such that the first \( p \) coordinates of \( Z_i \) are those with indices \( e_j, j = 1, \ldots, p \), where \( e_j \in \mathbb{Z}^p \) has the \( j \)-th coordinate equal to one and the rest equal to zero. Some of the coordinates of \( Z_i \) may be linearly dependent and we therefore write

\[
Z_i = \tilde{Z}_i A,
\]

where \( \tilde{Z}_i \) is of dimension \( l \leq k \), \( \tilde{Z}_i \) has linearly independent coordinates of which the first \( l_1 \) are continuous variables and the remaining \( l_2 = l - l_1 \) are lattice variables with minimal lattice \( \mathbb{Z}^2 \). We will write \( \tilde{Z}_i = (\tilde{Z}_i^{(1)}, \tilde{Z}_i^{(2)}) \), where \( \tilde{Z}_i^{(1)} \) are the first \( l_1 \) coordinates and \( \tilde{Z}_i^{(2)} \) the last \( l_2 \) coordinates.

We can now introduce the remaining assumptions needed. The first condition, apart from the moment conditions, is sometimes called a uniform Cramér condition, and is needed to establish an Edgeworth expansion for the continuous part given the lattice part. The second condition (B2) is intuitively the essential one. It is used to assure that a first order Taylor approximation of the statistic of interest depends on the continuous part.

(B1) \( E_{\theta_0} |\tilde{Z}_i^{(1)}|^{max \{2s+1, l_1+1\}} < \infty \), \( E_{\theta_0} |\tilde{Z}_i^{(2)}|^{max \{2s+1, l_1+1, l_2+1\}} < \infty \) and for all \( \epsilon > 0 \) there exists a \( \rho < 1 \) such that

\[
|E_{\theta_0} \exp \left( i \cdot \tilde{Z}_i^{(1)} + i \epsilon \cdot \tilde{Z}_i^{(2)} \right)| \leq \rho
\]

for \( |\epsilon| \leq \pi, j = 1, \ldots, l_2 \) and one of \( |\epsilon| > \epsilon \) or \( \epsilon > \epsilon \) being fulfilled.

(B2) The \( l_1 \times p \) matrix \( A^{(1)} \) has full rank \( p \), where \( A^{(1)} \) is the upper left hand corner of \( A \).

We write \( L_n(\theta) = \Sigma_{j=1}^{p} h(X_j; \theta) \) and consider the \( p \) equations

\[
1/n D_r L_n(\theta) = 0 \quad r = 1, \ldots, p.
\]

**Theorem 2.1**

(a) Assume (A1)–(A3) hold. Then there exists a sequence of statistics \( \{\hat{\theta}_n\}_{n \geq 1} \) and constants \( a_2, a_3 \) such that

\[
P_{\theta_0}(|\hat{\theta}_n - \theta_0| < a_2 (\log n/n)^{1/2}, \hat{\theta}_n \text{ solves (2.2)}) \geq 1 - a_4 n^{-(x-2)/2} (\log n)^{-s/2}.
\]

(b) Assume that (A1)–(A3), (B1) and (B2) hold. Then there exist polynomials \( q_{r, \theta_0} \) on \( \mathbb{R}^n \), not depending on \( n \), such that for any sequence satisfying (2.3) we have

\[
\left| P_{\theta_0}(n^{1/2}(\hat{\theta}_n - \theta_0) \in C) - \int_C \left[ 1 + \sum_{r=1}^{n^{1/2}} q_{r, \theta_0}(x) \right] \varphi_{\Sigma}(x) \, dx \right| = O(n^{-(x-2)/2})
\]

uniformly over the set of convex sets \( C \). Here \( \Sigma = I(\theta_0)^{-1} D(\theta_0) I(\theta_0)^{-1} \), and \( \varphi_{\Sigma} \) is the density of the multivariate normal distribution with variance \( \Sigma \).

*Proof.* The proof is almost identical to the proof of theorem 3 of Bhattacharya & Ghosh (1978), except for the use of theorem 1 of Jensen (1989b). We only sketch a few details needed in later proofs and the difference to Bhattacharya & Ghosh (1978).
Taylor expanding (2.2) to order $s$ and denoting the remainder term by $R_{n,s}(\theta)$, we have for some constant $d_i$

$$|R_{n,s}(\theta)| \leq d_i |\theta - \theta_0|^s \max_{|s| \leq 1} \sup_{|\theta - \theta_0| \leq \theta_0} |D' L_n(\theta)|.$$  

(2.4)

From assumption (A2) and corollary 17.12 in Bhattacharya & Rao (1976) we get the existence of constants $d_1$, $d_2$ and $d_4$ such that

$$P_{\theta_0} \left( \sqrt{n} Z^{(i)} - E_{\theta_0} Z^{(i)} \right) > d_2 (\log n)^{1/2} \leq d_3 n^{-(s-2)/2} (\log n)^{-s/2}$$  

(2.5)

for all $v$ with $|v| \leq s$, and

$$P_{\theta_0} \left( \max_{|s| \leq 1} \sup_{|\theta - \theta_0| \leq \theta_0} \left| \sum_{n=1} D' h(X_j; \theta) \right| \leq d_4 \right) \geq 1 - d_1 n^{-(s-2)/2} (\log n)^{-s/2}.$$  

(2.6)

Using the Brouwer fixed point theorem and the implicit function theorem it is shown that with probability at least $1 - d_3 n^{-(s-2)/2} (\log n)^{-s/2}$ there exists a solution $\tilde{\theta}_n$ to (2.2) with

$$|\tilde{\theta}_n - \theta_0| < d_4 (\log n/n)^{1/2}, \quad |R_{n,s}(\tilde{\theta}_n)| \leq d_4 (\log n/n)^{s/2},$$  

(2.7)

and there exists a smooth function $H$ defined on a neighbourhood of $\mu = (EZ^{(i)})_{i \leq |v| \leq s}$ in $R^k$ with $H(\mu) = \theta_0$ such that

$$\tilde{\theta}_n = H(Z') \quad \text{and} \quad |H(Z') - H(\tilde{\theta}_n)| \leq d_4 |R_{n,s}(\tilde{\theta}_n)|,$$  

(2.8)

where $Z^{(i)} = Z_{i}^{(i)}$ for $|v| \geq 2$ and $Z_{1}^{(i)} = Z_{1}^{(i)} + R_{n,s}(\tilde{\theta}_n)$.  

The remaining part of the proof now consists in establishing an Edgeworth expansion for the distribution of $\sqrt{n} [H(Z) - H(\mu)]$. This is done from theorem 1 of Jensen (1989b) using assumptions (B1) and (B2). Writing $\tilde{H}(\tilde{Z}) = H(\tilde{Z})$ assumption (B1) gives the necessary moment conditions and the uniform Cramér condition. Assumption (B2) is needed since from the equations $P(\tilde{H}(\tilde{Z}), \tilde{Z}; r) = 0, 1 \leq r \leq p$, (see Bhattacharya & Ghosh, 1978, equation 2.37) one finds that

$$\frac{\partial \tilde{H}(\mu)}{\partial Z^{(i)*}} = A^{(i)} \begin{pmatrix} I_p \\ 0 \end{pmatrix} f(\theta_0)^{-1} = A^{(i)1} f(\theta_0)^{-1},$$


which must have full rank. Here $A^{(i)}$ consists of the first $l_i$ rows of $A$ and $I_p$ is the $p \times p$ identity matrix. Finally, to transform the Edgeworth expansion for the distribution of $\sqrt{n} [H(Z) - H(\mu)]$ to $\sqrt{n} (\tilde{\theta}_n - \theta_0)$ we use (2.7) and (2.8). To perform this last step we need to make a restriction to convex sets only since the expansion of Jensen (1989b) is not for all Borel sets.

\textbf{Remark 2.2.} The difference to the result of Bhattacharya & Ghosh (1978) is that in their case $Z_{1} + \ldots + Z_{k}$ has an absolutely continuous component with respect to Lebesgue measure. Then the Edgeworth expansion for $\sqrt{n} [Hg(Z) - H(\mu)]$ holds uniformly for all Borel sets, and the resulting expansion for $\sqrt{n}(\tilde{\theta}_n - \theta_0)$ holds uniformly for classes of Borel sets with certain restrictions on their boundaries.

The uniform Cramér condition in (B1) can typically be proved from the following proposition.

\textbf{Proposition 2.3}

Let $(X, Y) \in \mathbb{R}^{l_1} \times \mathbb{R}^{l_2}$ and assume that the random vector has a continuous component with density $f(x, y)$ with respect to the product measure of Lebesgue measure and counting measure. Assume that there exist $c > 0$ and $a > 0$ such that

$$f(x, y) > c \quad \text{for} \quad (x, y) \in (-a, a)^{l_1} \times \{0, 1\}^{l_2}.$$
Then for all \( \varepsilon > 0 \) there exists \( \rho < 1 \) such that
\[
|E \exp(it \cdot X + iv \cdot Y)| < \rho
\]
for \( |v_j| \leq \pi, j = 1, \ldots, l_2 \) and one of \(|t| > \varepsilon \) or \(|v| > \varepsilon \) being fulfilled.

Proof. Let \( I_1 = (-\alpha, \alpha)^h \) and \( I_2 = \{0, 1\}^{l_2} \). The result follows trivially from the estimate
\[
|E \exp(it \cdot X + iv \cdot Y)| \leq \left| \sum_{y \in I_2 \setminus I_1} \int_{I_1} \exp(it \cdot x + iv \cdot y) c \, dx \right| + \sum_{y} \left| \int [f(x, y) - c_{1_{I_1} \times I_2}(x, y)] \, dx \right|
\]
\[
\leq c 2^{h(2a)^h} \left\{ \prod_{j=1}^{l_1} \int_{-a}^{a} \exp(it_j x_j) \frac{1}{2a} \, dx_j \right\} \left\{ \prod_{j=1}^{l_2} \frac{1 + \exp(iv_j)}{2} \right\}
\]
\[+ [1 - c 2^{h(2a)^h}]. \]

2.2. Edgeworth expansions for test statistics

We now turn to log likelihood ratio statistics. We use the set up from before with \( h(x; \theta) \) the logarithm of the density of \( P_\theta \) w.r.t. some measure \( \mu \), and let the true parameter value be \( \theta_0 = 0 \). We shall use the assumptions (A1) and (A2) with \( s = 4 \), and assume to begin with that the Fisher information \( I(\theta_0) \) is the \( p \times p \) identity matrix. Define
\[
l(\theta) = \frac{1}{n} \sum_{i=1}^{n} h(X_i; \theta),
\]
and let \( \bar{\theta} = (\bar{\theta}_1, \ldots, \bar{\theta}_p) \) and \( \bar{\theta} = (\bar{\theta}_1, \ldots, \bar{\theta}_{p-1}, 0) \) be the maximum likelihood estimates in the full model and in the restricted model with \( \theta_p = 0 \), respectively. Then the log likelihood ratio test statistics is
\[
W_n = 2n[l(\bar{\theta}) - l(\bar{\theta})].
\]

We first rewrite \( W_n \) in the form
\[
W_n = \sqrt{n[p - \bar{\theta}]J[\sqrt{n[p - \bar{\theta}]}}^* (2.9)
\]
with
\[
J = 2 \int_0^1 \int_0^u \frac{\partial^2 l}{\partial \bar{\theta} \partial \bar{\theta}^*} (\bar{\theta} + v(\bar{\theta} - \bar{\theta})) \, dv \, du.
\]
The matrix \( J \) will be close to the identity matrix since we have assumed (A2) and that the information matrix is the identity matrix. From \((\partial l/\partial \theta^{(1)}) (\bar{\theta}) = 0 \) and \((\partial l/\partial \theta^{(1)}) (\bar{\theta}) = 0 \), where \( \theta^{(1)} = (\theta_1, \ldots, \theta_{p-1}) \), we get
\[
0 = (\bar{\theta} - \bar{\theta}) \int_0^1 \frac{\partial^2 l}{\partial \theta^{(1)*} \partial \theta^{(1)}} (\bar{\theta} + v(\bar{\theta} - \bar{\theta})) \, dv = (\bar{\theta}^{(1)} - \bar{\theta}^{(1)}) \bar{J} - \bar{\theta} \bar{\delta},
\]
where
\[
\bar{J} = \int_0^1 \frac{\partial^2 l}{\partial \theta^{(1)*} \partial \theta^{(1)}} (\bar{\theta} + v(\bar{\theta} - \bar{\theta})) \, dv
\]
and
\[
\bar{\delta} = \int_0^1 \frac{\partial^2 l}{\partial \theta_p \partial \theta^{(1)}} (\bar{\theta} + v(\bar{\theta} - \bar{\theta})) \, dv.
\]
This gives

\[ \hat{\theta}^{(1)} - \hat{\theta}^{(0)} = \hat{\theta}^{(0)} \delta J^{-1}, \]

and using this in (2.9) we get

\[ W_n = (\sqrt{n}\hat{\theta})^2 \{ J_{22} + \delta J^{-1} J_{11}^{-1} \delta^* - 2\delta J^{-1} J_{12} \}, \]

where we have used obvious notation for the splitting of \( J \). Since \( J_{22} \) is close to 1 and \( \delta \) is close to zero we can write \( W_n = V_n^2 \) with

\[ V_n = \sqrt{n} \hat{\theta}_p \{ J_{22} + \delta J^{-1} J_{11}^{-1} \delta^* - 2\delta J^{-1} J_{12} \}^{1/2}. \]

Let now \( Z_i \) and \( \tilde{Z} \), be defined as in (2.1) and let \( U^{(i)} = \tilde{Z}^{(i)} - E\tilde{Z}^{(i)} \) for \( 1 \leq |v| \leq s \). It is then possible to show under (A1) and (A2), using (2.5), (2.6) and (2.7), that on a set having probability at least \( 1 - d_1 (1/n) (\log n)^{-2} \), \( V_n \) in (2.10) can be expanded as

\[ V_n = \sqrt{n} P_1(U) + \sqrt{n} P_2(U) + \sqrt{n} P_3(U) + \frac{1}{\sqrt{n}} R_n = \sqrt{n} P(U) + \frac{1}{\sqrt{n}} R_n \quad (2.11) \]

where

\[ P(|R_n| < d_2 (\log n/n)^2) \geq 1 - d_3 \frac{1}{n} (\log n)^{-2}. \quad (2.12) \]

Here \( P_i(U) \), \( i = 1, 2, \) and 3, are polynomials in the coordinates of \( U \) with all the terms of degree \( i \). The main term \( P_1(U) \) is

\[ P_1(U) = \frac{\partial l}{\partial \theta_p} (\theta_0) \quad (2.13) \]

expressed as a function of \( U \). If we can establish an Edgeworth expansion for \( \sqrt{n} P(U) \), we get from (2.11) and (2.12) an expansion for \( V_n \) which may then be transformed to an expansion for the distribution of \( W_n \).

In the general case where we want to test the hypothesis \( H: \theta_{1+1} = \ldots = \theta_p = 0 \), we let \( W_{nj} \) be the test statistic for testing \( \theta_j = 0 \) under the model with \( \theta_{j+1} = \ldots = \theta_p = 0 \) and write

\[ W_n = \sum_{j=p+1}^p W_{nj} = \sum_{j=p+1}^p V_{nj}^2 = \| V_n \|^2, \]

where \( V_{nj} \) has an expansion as in (2.11) and \( V_n = (V_{n,p+1}, \ldots, V_{n,p}) \). The statistic \( V_n \) has an expansion of the form

\[ V_n = \sqrt{n} P(U) + \text{remainder}, \quad (2.14) \]

where the main term of the expansion is

\[ \sqrt{n} \frac{\partial l}{\partial \theta_p} (\theta_0) = \sqrt{n}(Z_{p+1}, \ldots, Z_p). \]

We then get the following theorem by establishing an Edgeworth expansion for the distribution of \( \sqrt{n} P(U) \) from theorem 1 of Jensen (1989b).

**Theorem 2.4**

Assume (A1)–(A3) and (B1) with \( s = 4 \), that \( I(\theta_0) = I_p \), and that \( A^{(12)} \) has full rank, where \( A^{(12)} \) is the \( l_1 \times (p - p_1) \) matrix of the first \( l_1 \) rows and the columns \( (p_1 + 1, \ldots, p) \). Then there exists a polynomial \( q(v) \) such that

\[ \sup_u \left| P(W_n \leq u) - \int_0^u \left( 1 + \frac{1}{n} q(v) \right) f_{p-p_1}(v) \, dv \right| = O \left( \frac{1}{n} \right) , \]
where \( f_{p-p_1} \) is the chi-squared density on \( p - p_1 \) degrees of freedom. The polynomial \( q(v) \) is such that the expansion here is the same as that given in (1.9).

**Remark 2.5.** In the general case with \( l(\theta_0) = \Sigma \) positive definite we take \( \Sigma^{-1/2} \) such that if \( \vec{\theta} = \theta \Sigma^{-1/2} \) then \( \theta_{p_1+1} = \ldots = \theta_p = 0 \) corresponds to \( \vec{\theta}_{p_1+1} = \ldots = \vec{\theta}_p = 0 \). The condition \( A^{(12)} \) has full rank is then replaced by \((A^{(1)} \Sigma^{-1/2})^{(12)}\) has full rank, where \( A^{(1)} \) is the first \( p \) columns of \( A \).

**Remark 2.6.** When \( l = l_1 \), i.e. there are no lattice variables, the condition (B1) reduces to an ordinary Cramér condition on \( Z_l \), which then is an alternative to the assumption (A5) of section 1.1.

**Remark 2.7.** That the polynomial \( q(v) \) takes the same form as in the continuous case appears from Jensen (1989). In particular, as in theorem 1.1, a Bartlett adjustment exists such that the improved statistic \( WB_n \) has a chi-squared distribution up to terms of order \( O(1/n) \).

Let us now specialize the result of theorem 2.4 to an exponential family. Let

\[
\frac{dP_\theta}{d\mu}(\vec{z}) = \exp \{ \varphi(\theta) \cdot \vec{z} - \kappa(\varphi(\theta)) \},
\]

where \( \varphi(\theta) \) belongs to an open set of \( \mathbb{R}^l \) and \( \vec{Z} = (\vec{Z}^{(1)}, \vec{Z}^{(2)}) \) has linearly independent coordinates with the first \( l_1 \) having a continuous component w.r.t. Lebesgue measure and with the last \( l-l_1 \) coordinates having minimal lattice \( \mathbb{Z}^{l-l_1} \). There is no restriction in this formulation as long as we want the general formulation in (2.1) to hold, since the log likelihood derivatives here will be linear combinations of the coordinates of \( \vec{z} \). We want to test the hypothesis \( H: \theta_{p_1+1} = \ldots = \theta_p = 0 \). Let \( \theta_0 \) be the true parameter point, \( \varphi_0 = \varphi(\theta_0) \) and define

\[
\sum = \frac{\partial^2 \varphi}{\partial \theta^* \partial \theta^*}(\theta_0) \frac{\partial^2 \kappa}{\partial \theta^* \partial \theta}(\theta_0) \frac{\partial \varphi^*}{\partial \theta^*}(\theta_0).
\]

As before let \( \Sigma^{-1/2} \) be chosen such that if \( \vec{\theta} = \theta \Sigma^{-1/2} \) then \( \theta_{p_1+1} = \ldots = \theta_p = 0 \) corresponds to \( \theta_{p_1+1} = \ldots = \theta_p = 0 \).

To fulfill the uniform Cramér condition in (B1) we will use the following assumption.

(C1) There exist constants \( c > 0, a > 0, b_j \in \mathbb{R} \) for \( j = 1, \ldots, l_1 \) and \( \vec{b}_j \in \mathbb{Z} \) for \( j = 1, \ldots, l-l_1 \), such that \( \mu \) has an absolutely continuous component w.r.t. the product of Lebesgue measure in \( \mathbb{R}^{l_1} \) and counting measure in \( \mathbb{Z}^{l-l_1} \) which is greater than \( c \) on the set \( \Pi_{j=1}^{l_1} (b_j - a, b_j + a) \times \Pi_{j=1}^{l-l_1} \{b_j, b_j + 1\} \).

**Corollary 2.8**

Assume that the \( l_1 \times (p-p_1) \) matrix

\[
\left( \frac{\partial \varphi^*}{\partial \theta^*}(\theta_0) \Sigma^{-1/2} \right)^{(12)}
\]

has full rank \( p-p_1 \), and assume that (C1) is valid. Then the result of theorem 2.4 holds.

**Proof.** The first \( l \) coordinates of the \( Z \) variable in (2.1) are given by \( \bar{z}(\partial \varphi^*/\partial \theta^*) (\theta_0) \), which then defines the relevant part of the \( A \) matrix, and the result follows from theorem 2.4 and proposition 2.3. \( \square \)
We end this section by considering Bartlett adjusted statistics $WB_n$ of the form

$$WB_n = W_n/(1 + B/n)$$

(2.15)

where $B$ is stochastic, namely a smooth function of the variables $Z(\cdot)$, $1 \leq |v| \leq 4$. From (2.14) we see that $WB_n = |VB_n|^2$ with

$$VB_n = \sqrt{n} \left\{ P(U) \left[ 1 - \frac{1}{2n} B(Z) \right] \right\} + \text{remainder}.$$

Theorem 1 of Jensen (1989b) is also applicable in this situation, and since the derivative of $P(u)[1 - (1/2)(1/n)B(z)]$ w.r.t. $\tilde{z}$ converges to the derivatives of $P(u)$ w.r.t. $\tilde{z}$, we need the same assumptions as in theorem 2.4.

**Theorem 2.9**

Let the assumptions be as in theorem 2.4 and assume also that $B(z)$ in (2.15) is a smooth function of $z$ in a neighbourhood of $EZ_1$. Then the conclusion of theorem 2.4 holds with $W_n$ replaced by $WB_n$ from (2.15).

**Remark 2.10.** We have in general that the two Edgeworth expansions, including terms of order $O(n^{-1})$, for $W_n/(1 + B(Z)/n)$ and $W_n/(1 + B(EZ_1)/n)$ will be equal. If therefore $B(EZ_1)$ equals the theoretical Bartlett adjustment, i.e. $EW_n = (p - p_1) \{ 1 + B(EZ_1)/n + O(n^{-1}) \}$, then $WB_n = W_n/(1 + B(Z)/n)$ is chi-squared distributed with an error of order $O(n^{-1})$.

2.3. Example

We will consider a logistic regression model with the specification

$$P(Y = 1|X = x) = 1 - P(Y = 0|X = x) = \frac{e^{x + \beta X}}{1 + e^{x + \beta x}}$$

(2.16)

and $X$ has a density $h(x)$ w.r.t. Lebesgue measure. The log likelihood function based on $n$ observation is

$$l(x, \beta) = \sum_{i=1}^{n} Y_i X_i + \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \log (1 + e^{x + \beta x})$$

and we want to test the hypothesis $\beta = 0$. In this set up we find that

$$Z_i = \left(Y_i - \frac{e^x}{1 + e^x}, X_i \left(Y_i - \frac{e^x}{1 + e^x}\right), \ldots\right)$$

and that $Z_i$ can be taken as

$$Z_i = \left(X_i \left(Y_i - \frac{e^x}{1 + e^x}\right), X_i, X_i^2, X_i^3, X_i^4, Y_i\right)$$

i.e. only $Y_i$ has a lattice distribution, the remaining variables having a continuous distribution. Since

$$\frac{\partial l}{\partial \beta}(x, 0) = \sum_{i=1}^{n} X_i \left(Y_i - \frac{e^x}{1 + e^x}\right)$$
depends on the continuous part of \( \hat{Z} \) we have that the full rank condition of theorem 2.4 and theorem 2.9 is fulfilled. Thus these theorems are applicable under suitable moment conditions on the density \( h(x) \).

In particular we have in mind using theorem 2.9 with \( B \) the Bartlett adjustment calculated from the conditional distribution of \((Y_1, \ldots, Y_n)\) given \((X_1, \ldots, X_n)\), the latter can be calculated from the general formula in Lawley (1956) and is given by

\[
B = \frac{1}{4}(4 - e^x - e^{-x}) \{ (s_2 - s_1^2)^{-2}(3s_2^2 - 4s_1s_3 + s_4) - 1 \} \\
+ \frac{5}{12}(e^x + e^{-x} - 2) \{ (s_2 - s_1^2)^{-3}(4s_2^3 - 3s_1^2s_2^2 + 4s_1^3s_3 - 6s_1s_2s_3 + s_4^2) - 1 \}
\]  

(2.17)

where \( s_j = (1/n) \Sigma_{i=1}^n x_i^j \). From remark 2.10 we then get that \( WB_n = W_n/(1 + B/n) \) is chi-squared distributed with an error term of order \( O(n^{-2}) \) if for example all moments of \( h(x) \) exist.

The intriguing thing here is that the conditional Bartlett adjustment is used because one would like to make conditional inference given \((X_1, \ldots, X_n)\). But conditionally the adjusted log likelihood ratio statistic \( WB_n \) has a discrete distribution, and the expansion in a chi-squared series would probably not be valid. However, the unconditional distribution of \( WB_n \) has an expansion as stated in theorem 2.9.

The result of a simulation study can be seen in Table 1. The distribution of \( X \) was taken to be normal with mean zero and variance one, and \( x \) was taken to be zero. The true significance levels for \( W_n \) and \( WB_n \) were investigated at the nominal 5%, 2.5% and 1% levels. As can be seen from Table 1 the distribution of \( WB_n \) is very well described by a chi-squared distribution. For the case with the nominal level equal to 0.05 the errors, i.e. the true level minus the nominal level, times \( n \) for \( W_n \) and \( n^2 \) for \( WB_n \), respectively are also given in Table 1. These numbers are in agreement with the error being of order \( O(n^{-1}) \) for \( W_n \) and \( O(n^{-2}) \) for \( WB_n \).

3. Algebraic calculations using REDUCE

In this section we describe a program in the language REDUCE (Rayna, 1987), which may be used for producing the algebraic expression for the Bartlett adjustment.

### Table 1. True significance level for \( W_n \) and \( WB_n \) in the logistic regression example (2.16) with \( X \sim N(0, 1) \) and \( x = 0 \)

<table>
<thead>
<tr>
<th>Nominal level in percent</th>
<th>Error times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n )</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>5.00</td>
<td>2.50</td>
</tr>
<tr>
<td>6.30</td>
<td>3.33</td>
</tr>
<tr>
<td>6.66</td>
<td>3.60</td>
</tr>
<tr>
<td>5.24</td>
<td>2.66</td>
</tr>
<tr>
<td>6.02</td>
<td>3.18</td>
</tr>
<tr>
<td>5.08</td>
<td>2.56</td>
</tr>
<tr>
<td>5.57</td>
<td>2.87</td>
</tr>
<tr>
<td>5.01</td>
<td>2.51</td>
</tr>
</tbody>
</table>

The values are based on \( 10^6 \) simulations for \( n \leq 30 \) and \( 3 \cdot 10^6 \) simulations for \( n = 50 \). The last two columns are for the case with the nominal level equal to 0.05.
Let \( l(\theta) \) be the log likelihood function for the parameter \( \theta = (\theta_1, \ldots, \theta_p) \) divided by the number of observations \( n \). Define

\[
d2(i, j) = E_0 \frac{\partial^2 l}{\partial \theta_i \partial \theta_j}(\theta_0)
\]

\[
d3(i, j, k) = E_0 \frac{\partial^3 l}{\partial \theta_i \partial \theta_j \partial \theta_k}(\theta_0)
\]

\[
d4(i, j, k, l) = E_0 \frac{\partial^4 l}{\partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l}(\theta_0)
\]

\[
d2d1(i, j, k) = \frac{\partial}{\partial \theta_k} E_0 \frac{\partial^2 l}{\partial \theta_i \partial \theta_j}(\theta)|_{\theta=\theta_0}
\]

\[
d2d2(i, j, k, l) = \frac{\partial^2}{\partial \theta_j \partial \theta_l} E_0 \frac{\partial^2 l}{\partial \theta_i \partial \theta_k}(\theta)|_{\theta=\theta_0}
\]

\[
d3d1(i, j, k, l) = \frac{\partial}{\partial \theta_l} E_0 \frac{\partial^3 l}{\partial \theta_i \partial \theta_j \partial \theta_k}(\theta)|_{\theta=\theta_0}
\]

where \( 1 \leq i, j, k, l \leq p \). Lawley (1956) gave a formula for \( \varepsilon_p \) in

\[
E_0 \{2n[\max_\theta l(\theta) - l(\theta_0)]\} = p + \frac{1}{n} \varepsilon_p + O\left(\frac{1}{n^2}\right), \tag{3.1}
\]

based on the above defined quantities. This formula for \( \varepsilon_p \) is evaluated in the procedure LAWLEY(p) in Table 2 below when the \( d \)-functions have been defined.

---

**Table 2.** REDUCE procedure that evaluates the term \( \varepsilon_p \) in (3.1) based on the formula in Lawley (1956)

```
Procedure LAWLEY(p);
begin
scalar a,b;
clear dd2,dd3;
matrix dd2,p,p$;
for i=1:p do for j=1:p do dd2(i,j):=d2(i,j)$;
din:=1/dd2$;
for i=1:p do for j=1:p do for k=1:p do for f=1:p
  sum(din(i,j)*d3(i,j,k,f)/(d4(i,j,k,f)/4 - d3d1(i,k,f) + d2d2(i,j,k,f)))$;
for i=1:p do for j=1:p do for k=1:p do for f=1:p do for g=1:p do for h=1:p
  sum(d3(i,j,f)*d3(i,j,g,h)/(d3(i,j,k,f)*d3(i,j,f,h)/6 + d3(i,j,k,f)*d3(i,j,g,h)/4 - d3(i,k,g)*d3d1(i,j,h,f) - d3(i,j,f)*d2d1(i,j,h,g) + d2d1(i,k,g)*d2d1(i,j,h,f) + d2d1(i,k,f)*d2d1(i,j,h,g))$;
return a-b$;
end;
```

The operators \( d2-d4 \) and the arrays \( d2d1, d2d2, d3d1 \) must be specified before calling the procedure.
Consider now the hypothesis \( \theta^{(2)} = \theta_0^{(2)} \), where \( \theta \) is divided as \((\theta^{(1)}, \theta^{(2)})\) with \( p_1 \) and \( p - p_1 \) coordinates, respectively. Then

\[
E_{\theta_0}\{2n[\max_\theta l(\theta) - \max_\theta l(\theta^{(1)}, \theta_0^{(2)})]\}
= E_{\theta_0}\{2n[\max_\theta l(\theta) - l(\theta_0)] - 2n[\max_\theta l(\theta^{(1)}, \theta_0^{(2)}) - l(\theta^{(1)}, \theta_0^{(2)})]\}
= p - p_1 + \frac{1}{n}(e_p - e_{p_1}) + O\left(\frac{1}{n^2}\right),
\]

and the Bartlett adjustment \( B = e_p - e_{p_1} \) is obtained in REDUCE by writing

\[
B = \text{LAWLEY}(p) - \text{LAWLEY}(p_1);
\]

In an exponential family setting we can derive the \( d \)-functions from the cumulant transform and the functional form of the canonical parameter. It is therefore easy to write a general procedure in Reduce, on top of the procedure Lawley above, where the user only has to supply the cumulant transform and the canonical parameter expressed through \((\theta_1, \ldots, \theta_p)\).

For calculating \( B \) in (2.17) for the logistic regression example the simplest way is to define the \( d \)-functions directly in terms of \( s(k) = \sum_i x_i^k \), \( 0 \leq k \leq 4 \), and then call Lawley twice as above.

For more routine application of the Bartlett adjustment it would seem important to implement the calculations in the major statistical packages. Hopefully this will happen in the near future.

Acknowledgement

I should like to thank the referees for their careful reading.

References


*Received February 1991, in final form October 1991*

J. L. Jensen, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus, DK-8000, Aarhus C, Denmark.