COMPARISON OF TEST STATISTICS VIA EXPECTED LENGTHS OF ASSOCIATED CONFIDENCE INTERVALS

Rahul Mukerjee
Indian Institute of Management
Post Box No 16757 Calcutta 700 027, India

Nancy Reid
Department of Statistics
University of Toronto
Toronto, Ontario Canada M5S 3G3

Abstract

With reference to a large class of test statistics, higher order asymptotics on expected lengths of associated confidence intervals are investigated in a possibly non-iid setting. The connection with Bartlett adjustability is also indicated.

Key words: Bartlett adjustability; likelihood ratio statistic; Rao's score statistic; Wald's statistic.

AMS Subject Classification: 62F25, 62E20.

1 Introduction

Over the last two decades, there has been considerable interest in asymptotic comparison of test statistics. Fairly comprehensive results are now available with regard to higher order power under contiguous alternatives - see Amari (1985), Ghosh (1991), Taniguchi (1991), Mukerjee (1993) and Rao and Mukerjee (1996), among others, for reviews and recent results. Because of well-known duality relations, these results shed light on the power properties of the associated confidence sets; cf Amari (1985, Section 6.5). In the context of confidence sets, however, the expected volume (or length, in the one- parameter case) is also an important criterion and not much work has as yet been reported on the asymptotic comparison of test statistics from this standpoint. The present paper attempts to fill this gap to some extent.

We work under the same setup as in Taniguchi (1991) and consider a very large class of test statistics covering the likelihood ratio (LR), Rao's score and Wald's statistics (Rao, 1973, pp. 415-420), in particular. As in Taniguchi (1991), we consider the one-parameter case but allow a possibly non-iid setting and thus our results are applicable to models useful in time series analysis. After presenting some preliminaries in Section 2, explicit formulae for confidence intervals and their expected lengths are derived in Sections 3 and 4 respectively. In Section 4, we also observe an optimum property of the LR statistic, indicate the link with Bartlett adjustability and note the similarity with an optimum property of the score test with respect to third order average power. Unlike comparison of higher order power under contiguous alternatives, however, the present results are non-local. Illustrative examples are given in Section 5.

Kumon and Amari (1983) also compared test statistics in the one- parameter case via expected lengths of the associated confidence intervals. However, we consider a possibly non-iid setup, make no assumption about curved exponentiality and compare the test statistics in their original forms without any modification for local unbiasedness. Because of the last reason, our results differ from theirs even in their setting. See Madansky (1989) or Mukerjee (1993) for discussion on the motivation for comparing test statistics in their original forms.

2 Preliminary computation

Let $X^{(n)} = (X_1, \ldots, X_n)'$, $n \geq 1$, be a collection of possibly vector-valued random variables with density $f_n(x^{(n)}; \theta)$, where the parameter θ belongs to an open subset of $\mathcal{R}^1$. We work under the assumptions of Taniguchi (1991). 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 the fourth order, of

$$Z_i(\theta) = c_n^{-1} \left[\frac{d^i}{d\theta^i} \log f_n(X^{(n)}; \theta) - E_\theta \left\{ \frac{d^i}{d\theta^i} \log f_n(X^{(n)}; \theta) \right\} \right] \quad (i = 1, 2, 3) \tag{2.1}$$

possess asymptotic expansions of the form

$$\left. \begin{array}{l} \mathrm{cum}_\theta\{Z_i(\theta), Z_j(\theta)\} = k_{ij}^{(1)}(\theta) + c_n^{-2} k_{ij}^{(2)}(\theta) + o(c_n^{-2}) \ , \\ \mathrm{cum}_\theta\{Z_i(\theta), Z_j(\theta), Z_\ell(\theta)\} = c_n^{-1} k_{ij\ell}(\theta) + o(c_n^{-2}) \ , \\ \mathrm{cum}_\theta\{Z_i(\theta), Z_j(\theta), Z_\ell(\theta), Z_m(\theta)\} = c_n^{-2} k_{ij\ell m}(\theta) + o(c_n^{-2}) \ , \end{array} \right\} \tag{2.2}$$

where $k_{ij}^{(1)}(\theta)$, $k_{ij}^{(2)}(\theta)$, $k_{ij\ell}(\theta)$, $k_{ij\ell m}(\theta)$ are smooth functions which do not involve n. The validity of the third order Edgeworth expansion for the distribution of $Z_i(\theta), 1 \leq i \leq 3$, under θ, is also assumed. Let

$$\left.\begin{array}{l} I(\theta)=k_{11}^{(1)}(\theta),\ J(\theta)=k_{12}^{(1)}(\theta), L(\theta)=k_{13}^{(1)}(\theta),\ M(\theta)=k_{22}^{(1)}(\theta), \\ K(\theta)=k_{111}(\theta),\ N(\theta)=k_{112}(\theta), H(\theta)=k_{1111}(\theta),\ \Delta(\theta)=k_{11}^{(2)}(\theta), \\ \tilde{M}(\theta)=M(\theta)-\{J(\theta)\}^2/I(\theta), \quad \tilde{N}(\theta)=N(\theta)-J(\theta)K(\theta)/I(\theta), \\ W_1(\theta)=Z_1(\theta)\{I(\theta)\}^{-1/2}, \quad W_2(\theta)=Z_2(\theta)-\{J(\theta)/I(\theta)\}Z_1(\theta), \\ W_3(\theta)=Z_3(\theta)-\{L(\theta)/I(\theta)\}Z_1(\theta)\ . \end{array}\right\} \tag{2.3}$$

As in Taniguchi (1991) (see also Chandra and Mukerjee (1984)), we consider a class $\mathcal{F}$ of test statistics such that every statistic $T(\theta)$ in $\mathcal{F}$ admits an expansion of the form

$$\begin{aligned} T(\theta) &= \{W_1(\theta)\}^2+c_n^{-1}[a_1(\theta)\{W_1(\theta)\}^2W_2(\theta)+a_2(\theta)\{W_1(\theta)\}^3]+c_n^{-2}[b_1(\theta)\{W_1(\theta)\}^2 \\ &+ b_2(\theta)\{W_1(\theta)W_2(\theta)\}^2+b_3(\theta)\{W_1(\theta)\}^4+b_4(\theta)\{W_1(\theta)\}^3W_2(\theta)+b_5(\theta)\{W_1(\theta)\}^3W_3(\theta)] \\ &+ o_p(c_n^{-2}), \end{aligned} \tag{2.4}$$

under θ, where $a_1(\theta)$, $a_2(\theta)$ and $b_1(\theta),\cdots,b_5(\theta)$ are smooth nonrandom functions with functional forms free from n; these functions depend on the particular statistic $T(\theta)$ under consideration. Here we call $T(\theta)$ a statistic in the sense that the statistic $T(\theta_0)$ can be used for testing the hypothesis $\theta=\theta_0$. As observed in Taniguchi (1991), the class $\mathcal{F}$ is very natural and includes, in particular, the likelihood ratio statistic, Rao's score statistic, Wald's statistic $c_n^2 I(\hat{\theta})(\hat{\theta}-\theta)^2$ and a modified version of Wald's statistic with $I(\hat{\theta})$ replaced by $I(\theta)$; here $\hat{\theta}$ is the maximum likelihood estimator of θ based on $X^{(n)}$. For all these statistics

$$a_1(\theta)=\{I(\theta)\}^{-1}\beta_1, \qquad a_2(\theta)=\{I(\theta)\}^{-3/2}\{\beta_2 J(\theta)+\beta_3 K(\theta)\}, \tag{2.5}$$

where

$$\left.\begin{array}{ll} \beta_1=1,\beta_2=0,\beta_3=-1/3, & \text{for the LR statistic;} \\ \beta_1=\beta_2=\beta_3=0, & \text{for the score statistic;} \\ \beta_1=2,\beta_2=1,\beta_3=0, & \text{for Wald's statistic;} \\ \beta_1=2,\beta_2=\beta_3=-1, & \text{for modified Wald's statistic.} \end{array}\right\} \tag{2.6}$$

To facilitate the study of a confidence interval given by the inversion of $T(\theta)(\epsilon\mathcal{F})$, we now consider a "square root" version of $T(\theta)$ (cf. Bickel and Ghosh (1990)) and then a stochastic expansion thereof about $\hat{\theta}$. To that effect, from (2.4) note that

$$T(\theta)=\{T_{\text{root}}(\theta)\}^2+o_p(c_n^{-2}), \tag{2.7}$$

under θ, where

$$\begin{aligned} T_{\text{root}}(\theta) &= -W_1(\theta)+c_n^{-1}[A_1(\theta)W_1(\theta)W_2(\theta)+A_2(\theta)\{W_1(\theta)\}^2] \\ &+ c_n^{-2}[B_1(\theta)W_1(\theta)+B_2(\theta)W_1(\theta)\{W_2(\theta)\}^2+B_3(\theta)\{W_1(\theta)\}^3 \\ &+ B_4(\theta)\{W_1(\theta)\}^2W_2(\theta)+B_5(\theta)\{W_1(\theta)\}^2W_3(\theta)], \end{aligned} \tag{2.8}$$

$$\left.\begin{array}{l}A_1(\theta) = -\frac{1}{2}a_1(\theta),\ A_2(\theta) = -\frac{1}{2}a_2(\theta),\\ B_1(\theta) = -\frac{1}{2}b_1(\theta),\ B_2(\theta) = \frac{1}{8}\{a_1(\theta)\}^2 - \frac{1}{2}b_2(\theta),\ B_3(\theta) = \frac{1}{8}\{a_2(\theta)\}^2 - \frac{1}{2}b_3(\theta),\\ B_4(\theta) = \frac{1}{4}a_1(\theta)a_2(\theta) - \frac{1}{2}b_4(\theta), \quad B_5(\theta) = -\frac{1}{2}b_5(\theta).\end{array}\right\} \tag{2.9}$$

Next, to find an expansion for $T_{\rm root}(\theta)$ about $\hat\theta$, define for $1 \le i \le 4$,

$$\mu_i(\theta) = c_n^{-2} E_\theta \left\{ \frac{d^i}{d\theta^i} \log f_n(X^{(n)};\theta) \right\} . \tag{2.10}$$

By standard regularity conditions, from (2.2), (2.3) and (2.10),

$$\left.\begin{array}{l}\mu_1(\theta) = 0, \qquad \mu_2(\theta) = -\{I(\theta) + c_n^{-2}\Delta(\theta)\} + o(c_n^{-2}) ,\\ \mu_3(\theta) = -\{3J(\theta) + K(\theta)\} + o(c_n^{-1}),\\ \mu_4(\theta) = -\{4L(\theta) + 3M(\theta) + 6N(\theta) + H(\theta)\} + o(1) ,\end{array}\right\} \tag{2.11}$$

and

$$\left.\begin{array}{l}I'(\theta) = 2J(\theta) + K(\theta),\\ I''(\theta) = 2L(\theta) + 2M(\theta) + 5N(\theta) + H(\theta),\\ J'(\theta) = L(\theta) + M(\theta) + N(\theta), \quad K'(\theta) = 3N(\theta) + H(\theta),\end{array}\right\} \tag{2.12}$$

where primes denote differentiation with respect to θ. Also, by (2.1) and (2.10),

$$\left.\begin{array}{ll}Z_i'(\theta) = Z_{i+1}(\theta) + c_n\{\mu_{i+1}(\theta) - \mu_i'(\theta)\} & (i = 1,2,3),\\ Z_i''(\theta) = Z_{i+2}(\theta) + c_n\{\mu_{i+2}(\theta) - \mu_i''(\theta)\} & (i = 1,2),\\ Z_1'''(\theta) = Z_4(\theta) + c_n\mu_4(\theta), & \end{array}\right\} \tag{2.13}$$

where $Z_4(\theta)$ is defined as in (2.1).

Let $h = c_n(\theta - \hat\theta)$ and write $\hat Z_i = Z_i(\hat\theta)$, $\hat A_i = A_i(\hat\theta)$, $\hat B_i = B_i(\hat\theta)$, $\hat A_i' = A_i'(\hat\theta)$, $\hat I = I(\hat\theta)$, $\hat J = J(\hat\theta)$, $\hat I' = I'(\hat\theta)$, and so on. Since $\hat Z_1 = 0$ by the definition of $\hat\theta$, from (2.3) and (2.8) - (2.13), after a considerable algebra, one obtains

$$\begin{aligned} T_{\rm root}(\theta) = \hat I^{1/2}\{h \;&+\; c_n^{-1}(u_{11}h + u_{12}h^2)\\ &+\; c_n^{-2}(u_{21}h + u_{22}h^2 + u_{23}h^3)\} + o_p(c_n^{-2}), \end{aligned} \tag{2.14}$$

under θ, where

$$\begin{aligned} u_{11} &= -\hat Z_2(\hat A_1 + \hat I^{-1}), \quad u_{12} = \frac{1}{2}\hat I^{-1}\hat J + \hat I^{1/2}\hat A_2, \qquad (2.15)\\ u_{21} &= \hat I^{-1}(\hat\Delta + \hat Z_2^2\hat A_1) - \hat B_1 - \hat Z_2^2\hat B_2,\\ u_{22} &= \frac{1}{2}\hat I^{-2}(\hat I'\hat Z_2 - \hat I\hat Z_3) - (\hat Z_3 - \frac{1}{2}\hat I^{-1}\hat J\hat Z_2)\hat A_1 - \hat Z_2\hat A_1'\\ &\quad + \hat I^{1/2}(\hat Z_2\hat B_4 + \hat Z_3\hat B_5 - 2\hat I^{-1}\hat Z_2\hat A_2) ,\\ u_{23} &= \frac{1}{8}\hat I^{-2}\hat K(2\hat J + \hat K) + \frac{1}{12}\hat I^{-1}(2\hat L - 3\hat N - \hat H) - \hat I\hat B_3\\ &\quad - \frac{1}{2}\{\hat M + \hat N - \hat I^{-1}\hat J(\hat J + \hat K)\}\hat A_1 + \hat I^{1/2}(\hat A_2' + \hat I^{-1}\hat J\hat A_2) . \end{aligned}$$

Expansion (2.14) formally resembles the expansions for the posterior in Bayesian asymptotics: see, e.g. Ghosh and Mukerjee (1991). Note that the only place that θ occurs in expansion (2.14) is in h: this facilitates calculation of the endpoints of the confidence interval in the next section.

3 Confidence interval

For $i = 1, 2, \ldots$, let $G_i(\cdot)$ and $g_i(\cdot)$ denote respectively the cumulative distribution function and the probability density function of a central chi-square variate with i degrees of freedom. Note that for $t > 0$ and $i = 1, 2, \ldots$,

$$G_{i+2}(t) - G_i(t) = -2g_{i+2}(t), \quad g_{i+2}(t) = (t/i)g_i(t). \tag{3.1}$$

Using (3.1) in Theorem 1 of Taniguchi (1991) and proceeding as in Section 2 of Rao and Mukerjee (1996), after some simplification, one obtains

$$P_\theta[T(\theta) \leq z^2\{1 - c_n^{-2}\psi_T(\theta, z^2)\}] = 1 - \alpha + o(c_n^{-2}) \ , \tag{3.2}$$

where

$$\begin{aligned}
\psi_T(\theta, z^2) \ &= \ \frac{1}{2}(1 - z^2)\{I(\theta)\}^{-1}\tilde{N}(\theta)a_1(\theta) + \frac{1}{2}z^2(1 - \frac{1}{3}z^2)\{I(\theta)\}^{-3/2}K(\theta)a_2(\theta) \\
&+ \ \frac{3}{4}(1 - \frac{1}{3}z^2)\tilde{M}(\theta)\{a_1(\theta)\}^2 + \frac{1}{4}z^2(5 - z^2)\{a_2(\theta)\}^2 - b_1(\theta) \\
&- \ \tilde{M}(\theta)b_2(\theta) - z^2 b_3(\theta) - \{I(\theta)\}^{-1}\Delta(\theta) + \frac{1}{4}(1 - \frac{1}{3}z^2)\{I(\theta)\}^{-2}H(\theta) \\
&- \ \frac{5}{12}(1 - \frac{2}{3}z^2 + \frac{1}{15}z^4)\{I(\theta)\}^{-3}\{K(\theta)\}^2,
\end{aligned} \tag{3.3}$$

and z is the upper $\frac{1}{2}\alpha$-point of a standard normal variate.

Since

$$[z\{1 - \frac{1}{2}c_n^{-2}\psi_T(\hat{\theta}, z^2)\}]^2 = z^2\{1 - c_n^{-2}\psi_T(\theta, z^2)\} + o_p(c_n^{-2}) \ ,$$

by (2.7), (2.14) and (3.2), with error of approximation $o(c_n^{-2})$, a confidence set given by the inversion of $T(\theta)$ and having confidence coefficient $1 - \alpha$, can be expressed as

$$\begin{aligned}
S_T^*(X^{(n)}) = \{\theta \ \ &: \ \ |h + c_n^{-1}(u_{11}h + u_{12}h^2) + c_n^{-2}(u_{21}h + u_{22}h^2 + u_{23}h^3)| \\
&\leq \ \ \hat{z}(1 - \frac{1}{2}c_n^{-2}\hat{\psi}_T)\}
\end{aligned}$$

where $h = c_n(\theta - \hat{\theta})$, and

$$\hat{z} = z\hat{I}^{-1/2}, \qquad \hat{\psi}_T = \psi_T(\hat{\theta}, z^2) \ . \tag{3.4}$$

In fact, the above can be further approximated by the interval

$$
\begin{aligned}
S_T(X^{(n)}) = \{\theta : \; & - && \hat{z}(1 - \frac{1}{2}c_n^{-2}\hat{\psi}_T) - c_n^{-1}(-u_{11}\hat{z} + u_{12}\hat{z}^2) \\
& + && c_n^{-2}(-v_1\hat{z} + v_2\hat{z}^2 - v_3\hat{z}^3) \le c_n(\theta - \hat{\theta}) \\
& \le && \hat{z}(1 - \frac{1}{2}c_n^{-2}\hat{\psi}_T) - c_n^{-1}(u_{11}\hat{z} + u_{12}\hat{z}^2) \\
& + && c_n^{-2}(v_1\hat{z} + v_2\hat{z}^2 + v_3\hat{z}^3)\} \;, \qquad (3.5)
\end{aligned}
$$

where

$$
v_1 = u_{11}^2 - u_{21} \;, \qquad v_2 = 3u_{11}u_{12} - u_{22} \;, \qquad v_3 = 2u_{12}^2 - u_{23} \;, \qquad (3.6)
$$

in the sense that the P_θ-probability for the symmetric difference of $S_T(X^{(n)})$ and $S_T^*(X^{(n)})$ to include θ is of order $o(c_n^{-2})$. Denoting the length of the interval $S_T(X^{(n)})$ by s_T, by (3.4) and (3.5),

$$
c_n s_T = 2z\hat{I}^{-1/2}[1 - c_n^{-1}u_{11} + c_n^{-2}\{v_1 + v_3 z^2 \hat{I}^{-1} - \frac{1}{2}\psi_T(\hat{\theta}, z^2)\}] \;. \qquad (3.7)
$$

4 Expected length

We now proceed to find an expression for $c_n E_\theta(s_T)$. This calls for obtaining stochastic expansions about θ for the terms in the right-hand side of (3.7). For notational simplicity, hereafter we write $Z_i \equiv Z_i(\theta)$, $a_i \equiv a_i(\theta)$, $b_i \equiv b_i(\theta)$, $a_i' \equiv a_i'(\theta)$, $I \equiv I(\theta)$, $J \equiv J(\theta)$, $\tilde{M} \equiv \tilde{M}(\theta)$, $I' \equiv I'(\theta)$, and so on.

Since $\hat{\theta} = \theta - c_n^{-1}h$, by (2.9), (2.15) and (2.11) - (2.13),

$$
-u_{11} = Q_1 + c_n^{-1}Q_2 + o_p(c_n^{-1}) \;, \qquad (4.1)
$$

under θ, where

$$
\begin{aligned}
Q_1 &= (Z_2 + hJ)(I^{-1} - \frac{1}{2}a_1) \;, && (4.2) \\
Q_2 &= h(Z_2 + hJ)(I^{-2}I' + \frac{1}{2}a_1') - h\{Z_3 + \frac{1}{2}h(2L + M + N)\}(I^{-1} - \frac{1}{2}a_1) \;. && (4.3)
\end{aligned}
$$

By (3.7) and (4.1),

$$
\begin{aligned}
c_n s_T \;&=\; 2zI^{-1/2}[1 + c_n^{-1}(\frac{1}{2}I^{-1}I'h + Q_1) + c_n^{-2}\{\frac{1}{2}I^{-1}I'hQ_1 + Q_2 \\
&+\; v_1 + v_3 z^2\hat{I}^{-1} - \frac{1}{2}\psi_T(\hat{\theta}, z^2) \\
&+\; (\frac{3}{8}(I')^2 - \frac{1}{4}II'')I^{-2}h^2\}] + o_p(c_n^{-2}), \qquad (4.4)
\end{aligned}
$$

under θ. Now, from Taniguchi's (1991) equation (3.1),

$$h = -I^{-1}Z_1 + c_n^{-1}I^{-3}\{\frac{1}{2}(3J+K)Z_1^2 - IZ_1Z_2\} + o_p(c_n^{-1}) \ , \tag{4.5}$$

under θ. From (2.2), (2.3), (2.12), (3.3), (3.6) and (4.2) - (4.5), after tedious algebra, it follows that

$$c_n E_\theta(s_T) = 2zI^{-1/2}[1 + c_n^{-2}\{R(\theta,z) + V_T(\theta,z)\}] + o(c_n^{-2}) \ , \tag{4.6}$$

where $R(\theta, z)$, free from n, does not depend on a_1, a_2, $b_1, \ldots, b_5$ and hence is the same for all statistics in $\mathcal{F}$, and

$$\begin{aligned} V_T(\theta,z) &= \frac{1}{8}z^2\tilde{M}a_1^2 - \frac{1}{4}(1+z^2)I^{-1}\tilde{M}a_1 + \frac{1}{8}z^4a_2^2 \\ &- \frac{1}{4}z^2I^{-3/2}\{2J + (1-\frac{1}{3}z^2)K\}a_2 + \frac{1}{2}z^2I^{-1/2}a_2' \ . \end{aligned} \tag{4.7}$$

Remark 4.1 Equations (4.6) and (4.7) represent the main result of this paper. The statistics in $\mathcal{F}$ can be compared, with respect to expected lengths of the associated confidence intervals, on the basis of $V_T(\theta, z)$. By (4.7), it may be possible to discriminate among the members of $\mathcal{F}$ even in models where the quantity $\tilde{M}$, interpretable in terms of Efron's (1975) statistical curvature, vanishes (see Example 5.4 below); this is a consequence of comparing the statistics in $\mathcal{F}$ in their original forms without any correction for local unbiasedness. Interestingly, $V_T(\theta, z)$ does not depend on $b_1, \ldots, b_5$ and this is analogous to what happens while comparing the members of $\mathcal{F}$ with regard to higher order power – vide Rao and Mukerjee (1996).

Remark 4.2 While describing the class of statistics $\mathcal{F}$ in (2.4), it was implicit that $a_1, a_2, b_1, \ldots, b_5$ do not depend on z. Hence by (4.7), in general, a choice of the pair (a_1, a_2), minimizing $V_T(\theta, z)$ uniformly in θ and z, is not available. However, $V_T(\theta, z)$ is an even polynomial of degree four in z and its leading term, i.e. the coefficient of z^4, is minimized if and only if

$$a_2 = -\frac{1}{3}I^{-3/2}K \ , \tag{4.8}$$

a condition which holds for the LR statistic by (2.5) and (2.6). Thus, for large z (or equivalently, small α), the LR statistic will perform better, under the present criterion, than any other statistic in $\mathcal{F}$ for which (4.8) does not hold. As illustrated in the next section, for a given model, the explicit formula (4.7) helps in understanding this phenomenon in more detail. Note that the above property of the LR statistic resembles an optimum property of the score test, with respect to third order local average power, for small α – see Rao and Mukerjee (1996).

Remark 4.3 It is interesting to note that the condition (4.8) is also necessary for the Bartlett adjustability of $T(\theta)$ – see e.g., the proof of Theorem 2 in Taniguchi (1991). Thus, in a sense, the Bartlett adjustability of $T(\theta)$ has an influence on the expected length of the associated confidence interval.

5 Examples

We now consider the LR, Rao's score, Wald's and modified Wald's statistics in some detail with reference to certain specific models. By (2.5), (2.12) and (4.7), for these statistics, the expression for $V_T(\theta, z)$ reduces to

$$\begin{aligned} V_T(\theta,z) &= \frac{1}{8}\beta_1\{z^2(\beta_1-2)-2\}I^{-2}\tilde{M} \\ &+ \frac{1}{8}z^2[z^2\{\beta_2 J+(\beta_3+\frac{2}{3})K\}-8(2J+K)]I^{-3}(\beta_2 J+\beta_3 K) \\ &+ \frac{1}{2}z^2 I^{-2}\{\beta_2(L+M+N)+\beta_3(3N+H)\}\ , \end{aligned} \tag{5.1}$$

where β_1, β_2, β_3 are as in (2.6). The models in the first two examples relate to special cases of the Gaussian autoregressive moving average process; cf. Kendall, Stuart and Ord (1983, p. 527).

Example 5.1 Let $X^{(n)}$ represent a stretch of a Gaussian autoregressive process of order 1 with mean zero and spectral density

$$\phi_\theta(\lambda) = \sigma^2/\{2\pi(1-2\theta\cos\lambda+\theta^2)\}\ , \qquad \lambda\epsilon[-\pi,\pi]\ ,$$

where $\sigma^2(>0)$ is known and $\theta(|\theta|<1)$ is an unknown parameter. With $c_n = n^{1/2}$, following Taniguchi (1986), here

$$\begin{aligned} &I = 1/(1-\theta^2)\ , \quad K=-3J=6\theta/(1-\theta^2)^2\ , \quad L=0\ , \quad M=2(1+\theta^2)/(1-\theta^2)^3\ , \\ &N=-4(1+2\theta^2)/(1-\theta^2)^3\ , \quad H=6(3+7\theta^2)/(1-\theta^2)^3\ , \quad \tilde{M}=2/(1-\theta^2)^2\ . \end{aligned}$$

Hence by (2.6) and (5.1), for the LR, score, Wald's and modified Wald's statistics, the expression for $V_T(\theta, z)$ reduces respectively to

$$\begin{aligned} &V_{\mathrm{LR}}(\theta,z) = -\{\tfrac{1}{2}+\tfrac{5}{4}z^2+\tfrac{1}{2}z^4\theta^2/(1-\theta^2)\}, \quad V_{\mathrm{score}}(\theta,z)=0, \\ &V_{\mathrm{Wald}}(\theta,z) = -\{1+z^2+\tfrac{1}{2}z^4\theta^2/(1-\theta^2)\}, \quad V_{\mathrm{m\,Wald}}(\theta,z) = -(1+2z^2)\ . \end{aligned}$$

Thus $V_{\mathrm{LR}}(\theta,z) < V_{\mathrm{score}}(\theta,z)$ uniformly in θ and z. Also, $V_{\mathrm{LR}}(\theta,z) < V_{\mathrm{Wald}}(\theta,z)$ uniformly in θ if and only if $z^2 > 2$, a condition which holds for $\alpha = 0.10$, 0.05, 0.025 and 0.01. Finally, $V_{\mathrm{LR}}(\theta,z) < V_{\mathrm{m\,Wald}}(\theta,z)$ if and only if

$$\theta^2 > (2+3z^2)/(2+3z^2+2z^4)\ .$$

Hence for $\alpha = 0.10$, 0.05, 0.025 and 0.01, $V_{\mathrm{LR}}(\theta, z)$ is less than $V_{\mathrm{m\,Wald}}(\theta, z)$ if and only if $|\theta|$ exceeds 0.639, 0.560, 0.502 and 0.446 respectively, i.e., as anticipated from Remark 4.2, with smaller α the performance of the LR statistic improves.

The explicit formulae (4.7) or (5.1) also allow comparison of statistics via an integrated average of $V_T(\theta, z)$ with respect to some weight function in θ. This may help when, for fixed z, no uniform result in θ is available. Thus, in this example, considering the weight function $w(\theta) = (2/\pi)(1-\theta^2)^{1/2}$, $|\theta| < 1$,

$$\int_{-1}^{1} V_{\mathrm{LR}}(\theta, z) w(\theta) d\theta = -\tfrac{1}{4}(2 + 5z^2 + 2z^4) ,$$
$$\int_{-1}^{1} V_{\mathrm{m\,Wald}}(\theta, z) w(\theta) d\theta = -(1 + 2z^2) ,$$

and the former is less than the latter if and only if $z^2 > 2$ (i.e., for $\alpha = 0.10$, 0.05, 0.025 and 0.01).

Example 5.2 Let $X^{(n)}$ represent a stretch of a Gaussian moving average process of order 1 with mean zero and spectral density

$$\phi_\theta^*(\lambda) = \{\sigma^2/(2\pi)\}(1 - 2\theta\cos\lambda + \theta^2) , \qquad \lambda\epsilon[-\pi, \pi] ,$$

where $\sigma^2 (> 0)$ is known and $\theta(|\theta| < 1)$ is an unknown parameter. With $c_n = n^{1/2}$, following Taniguchi (1986), then

$$I = 1/(1-\theta^2) , \qquad J = -\tfrac{2}{3}K = 4\theta/(1-\theta^2)^2 , \quad L = 6(1+3\theta^2)/(1-\theta^2)^3 ,$$
$$M = \tfrac{1}{3}H = 2(3+7\theta^2)/(1-\theta^2)^3 , \quad N = -4(2+5\theta^2)/(1-\theta^2)^3 ,$$

so that $\tilde{M} = 2(3-\theta^2)/(1-\theta^2)^3$, and, as in Example 5.1, by (2.6) and (5.1),

$$V_{\mathrm{LR}}(\theta, z) = -\tfrac{1}{4}\{2(3-\theta^2) - z^2(1-3\theta^2) + 2z^4\theta^2\}/(1-\theta^2) ,$$
$$V_{\mathrm{score}}(\theta, z) = 0 , \qquad V_{\mathrm{Wald}}(\theta, z) = 2z^2 - (3-\theta^2)/(1-\theta^2) ,$$
$$V_{\mathrm{m\,Wald}}(\theta, z) = z^2 - (3 - \theta^2 + \tfrac{1}{2}z^4\theta^2)/(1-\theta^2) .$$

It can be seen that for $\alpha = 0.10$, 0.05, 0.025 and 0.01, $V_{\mathrm{LR}}(\theta, z)$ is less than both $V_{\mathrm{Wald}}(\theta, z)$ and $V_{\mathrm{m\,Wald}}(\theta, z)$ uniformly in θ. Also, $V_{\mathrm{LR}}(\theta, z) < V_{\mathrm{score}}(\theta, z)$ if and only if

$$6 - z^2 + (2z^4 + 3z^2 - 2)\theta^2 > 0 , \tag{5.2}$$

a condition which holds for every θ when $\alpha = 0.10$, 0.05 and 0.025 and for $|\theta| > 0.077$ when $\alpha = 0.01$. Incidentally, the last observation is not in conflict with Remark 4.2 since for each $\theta \neq 0$, the inequality (5.2) holds, indeed, for sufficiently large z.

Example 5.3 Let $X_1, X_2, \ldots$ be iid each univariate normal with mean θ and variance θ^2. Then

$$I = 3/\theta^2 , \quad J = -10/\theta^3 , \quad K = 14/\theta^3 , \quad L = 42/\theta^4 , \quad M = 34/\theta^4 ,$$
$$N = -46/\theta^4 , \quad H = 96/\theta^4 , \quad \tilde{M} = 2/(3\theta^4) ,$$

and, as before,

$$\begin{array}{ll} V_{\rm LR}(\theta,z) = -\frac{1}{972}(18+261z^2+98z^4)\ , & V_{\rm score}(\theta,z)=0\ , \\ V_{\rm Wald}(\theta,z) = -\frac{1}{162}(6+90z^2-5z^4)\ , & V_{\rm m\,Wald}(\theta,z) = -\frac{1}{81}(3+18z^2+8z^4)\ . \end{array}$$

Hence for $\alpha = 0.10,\ 0.05,\ 0.025$ and 0.01, $V_{\rm LR}(\theta,z)$ is less than each of $V_{\rm score}(\theta,z)$, $V_{\rm Wald}(\theta,z)$ and $V_{\rm m\,Wald}(\theta,z)$ uniformly in θ.

Example 5.4 Let $X_1, X_2, \ldots$ be iid each univariate normal with mean zero and variance θ. Then

$$I = 1/(2\theta^2)\ , \quad K=-J=1/\theta^3\ , \quad L=H=3/\theta^4\ , \quad M=-N=2/\theta^4\ .$$

Hence $\tilde{M}=0$ and

$$\begin{array}{ll} V_{\rm LR}(\theta,z) = -\frac{1}{9}z^2(6+z^2)\ , & V_{\rm score}(\theta,z)=0\ , \\ V_{\rm Wald}(\theta,z) = \frac{1}{3}z^2(z^2-6)\ , & V_{\rm m\,Wald}(\theta,z)=0\ . \end{array}$$

Thus $V_{\rm LR}(\theta,z)$ is less than both $V_{\rm score}(\theta,z)$ and $V_{\rm m\,Wald}(\theta,z)$ uniformly in θ and z. Also, $V_{\rm LR}(\theta,z) < V_{\rm Wald}(\theta,z)$, uniformly in θ, for $\alpha = 0.05,\ 0.025$ and 0.01.

Remark 5.1 The above examples lend support to Remark 4.2 and demonstrate that the LR statistic tends to perform better than the others for small but reasonable values of α. It may also be of interest to consider a modified version of the score statistic which is based on the "observed" rather than "expected" information and defined as $T_{\rm m\,score}(\theta) = \{\hat{I}(\theta)\}^{-1}\{Z_1(\theta)\}^2$, where

$$\hat{I}(\theta) = -c_n^{-2}\frac{d^2}{d\theta^2}\log f_n(X^{(n)};\theta)\ .$$

This also belongs to the class $\mathcal{F}$, with $a_1 = I^{-1}$ and $a_2 = I^{-3/2}J$. Revisiting the examples, one can check that the modification improves upon the performance of the score statistic in some models (e.g., in Examples 5.1, 5.3 and 5.4) but continues to be dominated by the LR statistic in most situations.

Remark 5.2 It can be shown that in (4.6)

$$\begin{aligned} R(\theta,z) &= \frac{1}{24}I^{-3}\{36J^2+20K^2+66JK-12I^2\Delta-3I(4L+14N+3H)\} \\ &\quad + \frac{1}{72}z^2I^{-3}\{36J^2-19K^2-18JK+3I(6N+3H-4L)\} \\ &\quad + \frac{1}{72}z^4I^{-3}K^2. \end{aligned}$$

One may be interested in comparing the variation in $V_T(\theta,z)$ over various choices of the statistics $T(\theta)$ with the magnitude of $R(\theta,z)$ which is the same for all statistics. Numerical studies reveal that such variation can be substantial. For instance, in the setup of Example 5.3, $R(\theta, z) = (1/486)(248z^2 + 49z^4 - 21)$ and, with $\alpha = 0.01$, $R(\theta, z)$ equals 7.781 while $V_{\mathrm{LR}}(\theta, z)$, $V_{\mathrm{score}}(\theta, z)$, $V_{\mathrm{Wald}}(\theta, z)$ and $V_{\mathrm{m\,Wald}}(\theta, z)$ equal -6.239, 0, -2.364, and -5.859 respectively.

Acknowledgement: This work, done while RM was visiting the University of Toronto, was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Centre for Management and Development Studies, Indian Institute of Management Calcutta. Thanks are due to the referees for very constructive suggestions.

References

Amari, S. (1985). *Differential Geometrical Methods in Statistics.* Springer-Verlag, Berlin.

Bickel, P.J. and 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. and Mukerjee, R. (1984). On the optimality of Rao's statistic. *Commun. Statist. – Theory Meth.* **13**, 1507–1515.

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.* **3**, 1189-1242.

Ghosh, J.K. (1991). Higher order asymptotics for the likelihood ratio, Rao's and Wald's tests. *Statist. Probab. Lett.* **12**, 505-509.

Ghosh, J.K. and Mukerjee, R. (1991). Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case. *J. Mult. Anal.* **38**, 385-393.

Kendall, M., Stuart, A. and Ord, J.K. (1983). *The Advanced Theory of Statisics*, Volume 3 (4th ed.). Charles Griffin, London.

Kumon, M. and Amari, S. (1983). Geometrical theory of higher order asymptotics of test, interval estimator and conditional inference. *Proc. Roy. Soc. London A* **387**, 429-458.

Madansky, A. (1989). A comparison of the likelihood ratio, Wald and Rao tests. In: Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin (L.J. Glesar et al eds.), pp. 465-471, Springer-Verlag, Berlin.

Mukerjee, R. (1993). Rao's score test: Recent asymptotic results. In: *Handbook of Statistics Vol. 11* (G.S. Maddala et al eds), pp. 363-379, North-Holland, Amsterdam.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications (2nd ed.).* John Wiley, New York.

Rao, C.R. and Mukerjee, R. (1996). Comparison of LR, score and Wald tests in a non-iid setting. *J. Mult. Anal.*, to appear.

Taniguchi, M. (1986). Third-order asymptotic properties of maximum likelihood estimators for Gaussian ARMA processes. *J. Mult. Anal.* **18**, 1-31.

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