Statistica Sinica **10**(2000), 297-315

LOCAL COMPARISON OF RAO AND WALD STATISTICS IN THE BAHADUR SENSE

Yoshihide Kakizawa

Hokkaido University

Abstract: Global optimality of likelihood ratio test statistics is well-known in the Bahadur sense. In this paper the behaviors of Rao and Wald statistics (R_n and W_n) for testing $\theta = \theta_0$ are studied. It turns out that at alternative $\theta_0 + \varepsilon$, the Bahadur slopes of these two statistics for the one-sided case are identical up to order ε^4, while for the two-sided case, they are identical only up to order ε^2, in general i.i.d. models and Gaussian stationary processes. We obtain the second- (first-) order Bahadur efficiency of R_n and W_n for the one- (two-) sided case. The third-order Bahadur efficiency depends on the statistical curvature. Two concrete examples are given. One is a curved exponential family, and the other is a Gaussian AR(1) process. The latter provides an example that the ε^5-term of the Bahadur slope of R_n for the one-sided case is different from that of W_n.

Key words and phrases: Bahadur slope, curved exponential family, Gaussian stationary process, large deviation theorem, Rao's statistic, spectral density, Wald's statistic.

1. Introduction

Let $X, X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.) random variables (or vectors) with a family of densities $\{p(\cdot;\theta)\}$ with respect to a σ-finite measure on the Borel sets of an Euclidean space, where the parameter θ takes values in some interval Θ on the real line. Consider the problem of testing $H : \theta = \theta_0$, where θ_0 is given. Higher order comparison of tests under contiguous alternatives via the χ^2 type asymptotic expansion has received considerable attention (e.g. Mukerjee (1993) and Rao and Mukerjee (1997)). Our study is based on the large deviation theory approach. Using the concept of exact slope by Bahadur (1960), we examine the performance of Rao and (modified) Wald statistics in the testing problem $H : \theta = \theta_0$ against $A_1 : \theta > \theta_0$ or $A_2 : \theta \neq \theta_0$.

Let T_n be a test statistic based on n observations $X_1, \ldots, X_n$ such that large values of T_n are significant. For any t, let $F_n(t) = P_{\theta_0}(T_n < t)$. The level attained by T_n is given by $L_n(T_n) = 1 - F_n(T_n)$. In typical cases there exists $c(\theta)$ such that $0 < c(\theta) < \infty$ and

$$\lim_{n\to\infty} n^{-1} \log L_n(T_n) = -\frac{1}{2}\, c(\theta) \tag{1}$$

with probability one, or in probability, when θ obtains. In accordance with the terminology of Raghavachari (1970), we call $c(\theta)$ the strong Bahadur slope of T_n at θ if (1) happens with probability one and the weak Bahadur slope of T_n at θ if (1) happens in probability. For details regarding comparison of test statistics through the slopes, refer to Serfling (1980, 10.4). In what follows, the strong Bahadur slope will be referred to as the Bahadur slope.

It is well-known from Raghavachari (1970) or Bahadur (1971, p.29) that for each alternative θ,

$$\liminf_{n\to\infty} n^{-1}\log L_n(T_n) \geq -K(\theta,\theta_0) \text{ with } P_\theta \text{ probability one,} \tag{2}$$

where $K(\theta,\theta_0)$ is the Kullback-Leibler information given by

$$K(\theta,\theta_0) = \int \Big\{\log \frac{p(x;\theta)}{p(x;\theta_0)}\Big\} p(x;\theta)\, dx\,. \tag{3}$$

In other words, the slope $c(\theta)$ cannot exceed $2K(\theta,\theta_0)$. Bahadur (1971, p.37) proved that the likelihood ratio test statistic has the slope $2K(\theta,\theta_0)$ at all alternative θ under certain general conditions. Akritas and Kourouklis (1988) showed that Rao's statistic

$$R_n = n^{-1/2}\sum_{i=1}^n \frac{\partial}{\partial\theta}\log p(X_i;\theta)\Big|_{\theta=\theta_0} \quad \text{or} \quad n^{-1/2}\Big|\sum_{i=1}^n \frac{\partial}{\partial\theta}\log p(X_i;\theta)\Big|_{\theta=\theta_0}$$

is locally optimal in the sense that $c_R(\theta)/\{2K(\theta,\theta_0)\} \to 1$ as $\theta \to \theta_0$. He and Shao (1996) also considered local optimality of the studentized score test. In the estimation case, Fu (1982) discussed asymptotic efficiency of the maximum likelihood estimator (MLE) with respect to the exponential rate $\lim_{n\to\infty} n^{-1} \log P_{\theta_0}(|\widehat{\theta}_{n,\mathrm{ML}} - \theta_0| \geq \varepsilon)$. He derived the first four terms of the Taylor expansions of $\lim_{n\to\infty} n^{-1}\log P_{\theta_0}(\widehat{\theta}_{n,\mathrm{ML}} \geq \theta_0+\varepsilon)$ and $\lim_{n\to\infty} n^{-1}\log P_{\theta_0}(\widehat{\theta}_{n,\mathrm{ML}} \leq \theta_0-\varepsilon)$. Since, from the point of view of the testing theory, Fu's work leads to the first four terms of the Taylor expansion of the Bahadur slope of the Wald's statistic $W_n = \sqrt{n}(\widehat{\theta}_{n,\mathrm{ML}} - \theta_0)$ or $\sqrt{n}|\widehat{\theta}_{n,\mathrm{ML}} - \theta_0|$, the aim of this paper is to make a local comparison between R_n and W_n in the Bahadur sense, and to elucidate their local optimality.

Although Bahadur asymptotic efficiency of estimators was originally developed for the i.i.d. setting, Sato, Kakizawa and Taniguchi (1998) recently mentioned that the MLE of the spectral parameter in a Gaussian stationary process is asymptotically Bahadur efficient. As a continuation of their work, Kakizawa (1997) did further study of the exponential rates of the MLE and the quasi-MLE of the spectral parameter. Thus we are interested in the performance of Rao's statistic for Gaussian stationary processes as well as for i.i.d. models.

The paper is organized as follows. In Section 2, we begin by deriving the first four terms of the Bahadur slope of R_n for general i.i.d. models and Gaussian stationary processes. By using the results on the exponential rate of the MLE (see Fu (1982) for the i.i.d. case and Kakizawa (1997) for Gaussian stationary processes), it is shown that at alternative $\theta_0+\varepsilon$, the Bahadur slopes of R_n and W_n for the one-sided case are identical up to order ε^4, while for the two-sided case, they are identical only up to order ε^2. This yields the second– (first–) order Bahadur efficiency of R_n and W_n for the one- (two-) sided case. Third order Bahadur efficiency depends on the statistical curvature. In Section 3, we give proofs of these results. The final section contains two concrete examples for illustrating the results for the one-sided case in Section 2. One is a curved exponential family, and the other is a Gaussian AR(1) process. The latter provides an example in which the ε^5-term of the Bahadur slope of R_n for the one-sided case is different from that of W_n.

2. Main Results

2.1. IID case

Let $X, X_1, X_2, \ldots$ be as in the introduction and write $\ell(x;\theta)=\log p(x;\theta)$. Assuming that for almost all x, $p(x;\theta)$ is four times continuously differentiable in θ, let $\ell^{(i)}(x;\theta)=(\partial/\partial\theta)^i \log p(x;\theta)$. We set down the following conditions.
(A1) For every $\theta_0\in\Theta$, there exists a constant $u=u(\theta_0)>0$ such that the moment generating function (MGF) $E_{\theta_0}[\exp\{t\ell^{(1)}(X;\theta_0)\}]=m(t)$ (say) is finite for $t\in(-u,u)$. We set $m(t)=+\infty$ for $t\notin(-u,u)$.

It follows that derivatives of all orders of $m(t)$ exist for $t\in(-u,u)$ and are given by differentiation under the integral sign, that is,

$$\frac{d^i}{dt^i}m(t)=E_{\theta_0}[\{\ell^{(1)}(X;\theta_0)\}^i\exp\{t\ell^{(1)}(X;\theta_0)\}]$$

for $i\geq 1$. We write $\mu_i=E_{\theta_0}[\{\ell^{(1)}(X;\theta_0)\}^i]$, $i=1,\ldots,4$.
(A2) For every $\theta_0\in\Theta$, $\ell^{(1)}(X;\theta_0)$ is non-degenerate.
(A3) For every $\theta_0\in\Theta$, there exist a neighborhood $U(\theta_0)$ and measurable functions $A_i(x,\theta_0)$ such that

a) $|\ell^{(i)}(x;\theta)|<A_i(x,\theta_0)$ for all $\theta\in U(\theta_0)$, $i=1,\ldots,4$, and

b) $E_{\theta_0}[\{A_i(X,\theta_0)\}^{k_i}]<\infty$, where $k_1=8, k_2=k_3=4$ and $k_4=2$.

Define $I(\theta),\ldots,H(\theta)$ by

$$\begin{aligned}
&I(\theta)=E_\theta[\{\ell^{(1)}(X;\theta)\}^2]\,,\\
&J(\theta)=E_\theta[\ell^{(1)}(X;\theta)\ell^{(2)}(X;\theta)]\,,\quad K(\theta)=E_\theta[\{\ell^{(1)}(X;\theta)\}^3]\,,\\
&L(\theta)=E_\theta[\ell^{(1)}(X;\theta)\ell^{(3)}(X;\theta)]\,,\quad M(\theta)=E_\theta[\{\ell^{(2)}(X;\theta)\}^2]-I^2(\theta)\,,\\
&N(\theta)=E_\theta[\{\ell^{(1)}(X;\theta)\}^2\ell^{(2)}(X;\theta)]+I^2(\theta)\,,\\
&H(\theta)=E_\theta[\{\ell^{(1)}(X;\theta)\}^4]-3I^2(\theta)\,.
\end{aligned}$$

To simplify the notation, we write $I(\theta_0), \ldots, H(\theta_0)$, as $I, \ldots, H$, respectively. These notations have been defined by Akahira and Takeuchi (1981, p.141) in order to discuss the higher order asymptotic theory of estimators. Note that by (A1) and (A2),

$$\mu_2 = I > 0\,, \mu_3 = K \text{ and } \mu_4 - 3\mu_2^2 = H\,. \tag{4}$$

By (A3), we have $\mu_1 = E_{\theta_0}[\ell^{(1)}(X;\theta_0)] = 0$, $E_{\theta_0}[\ell^{(2)}(X;\theta_0)] = -I$, $E_{\theta_0}[\ell^{(3)}(X;\theta_0)] = -3J - K$ and $E_{\theta_0}[\ell^{(4)}(X;\theta_0)] = -4L - 6N - 3M - H$. Further, the Kullback-Leibler information (3) has a four-term Taylor expansion

$$\begin{aligned} K(\theta_0+\varepsilon,\theta_0) &= E_{\theta_0}[\{\ell(X;\theta_0+\varepsilon) - \ell(X;\theta_0)\}\exp\{\ell(X;\theta_0+\varepsilon) - \ell(X;\theta_0)\}] \\ &= \frac{1}{2}\,I\varepsilon^2 + \frac{1}{6}\,(3J+2K)\varepsilon^3 + \frac{1}{24}\,(4L+3M+12N+3H)\varepsilon^4 + \mathrm{o}(\varepsilon^4)\,. \end{aligned} \tag{5}$$

We say that a test statistic T_n is sth order efficient in the Bahadur sense if the Bahadur slope c of T_n satisfies

$$\lim_{\varepsilon\to 0} \varepsilon^{-(s+1)}\,[2K(\theta_0+\varepsilon,\theta_0) - c(\theta_0+\varepsilon)] = 0\,.$$

Theorem 1. [One-sided case A_1]

(i) *The Bahadur slope of Rao's statistic*

$$R_n = n^{-1/2}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\,\log p(X_i;\theta)\Big|_{\theta=\theta_0}$$

at $\theta=\theta_0+\varepsilon$, where $\varepsilon>0$, has an expansion

$$\begin{aligned} c_R(\theta_0+\varepsilon) &= I\varepsilon^2 + \frac{1}{3}\,(3J+2K)\varepsilon^3 \\ &\quad + \frac{1}{12}\,\Big(3\frac{J^2}{I} + 4L + 12N + 3H\Big)\varepsilon^4 + \mathrm{o}(\varepsilon^4)\,, \end{aligned} \tag{6}$$

so that R_n is second-order efficient in the Bahadur sense.

(ii) *The difference between the optimal slope $2K(\theta_0+\varepsilon,\theta_0)$ and $c_R(\theta_0+\varepsilon)$ appears in the fourth-order term through Efron's statistical curvature*

$$\gamma(\theta) = \frac{\{M(\theta)I(\theta) - J(\theta)^2\}^{\frac{1}{2}}}{I(\theta)^{\frac{3}{2}}}\,,$$

that is,

$$\lim_{\varepsilon\to 0}\varepsilon^{-i}\,[2K(\theta_0+\varepsilon,\theta_0) - c_R(\theta_0+\varepsilon)] = \begin{cases} 0\,, & i = 1,2,3 \\ \dfrac{1}{4}\,I(\theta_0)^2\gamma(\theta_0)^2, & i = 4\,. \end{cases} \tag{7}$$

(iii) R_n *is locally equivalent to Wald's statistic* $W_n = \sqrt{n}(\widehat{\theta}_{n,\mathrm{ML}} - \theta_0)$ *up to order* ε^4 *in the sense that* $\lim_{\varepsilon\to 0}\varepsilon^{-4}\,[c_R(\theta_0+\varepsilon) - c_W(\theta_0+\varepsilon)] = 0$.

Theorem 2. [Two-sided case A_2]

(i) *The Bahadur slopes of Rao's statistic*

$$R_n = n^{-1/2}\left|\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log p(X_i;\theta)\right|_{\theta=\theta_0}$$

and Wald's statistic $W_n = \sqrt{n}|\widehat{\theta}_{n,\mathrm{ML}} - \theta_0|$ *at* $\theta = \theta_0 + \varepsilon$*, where* $\varepsilon \neq 0$*, have expansions*

$$c_R(\theta_0+\varepsilon) = \min\Big[\varepsilon^2 I + \frac{\varepsilon^3}{3}\,(3J+2K) + \mathrm{O}(\varepsilon^4)\,, \varepsilon^2 I + \frac{\varepsilon^3}{3}\,(3J+4K) + \mathrm{O}(\varepsilon^4)\Big]$$

and

$$c_W(\theta_0+\varepsilon) = \min\Big[\varepsilon^2 I \pm \frac{\varepsilon^3}{3}\,(3J+2K) + \mathrm{O}(\varepsilon^4)\Big]\,,$$

respectively. Both R_n *and* W_n *are only first-order efficient in the Bahadur sense.*

(ii) R_n *is second-order efficient in the Bahadur sense if and only if* $K(\theta_0) = 0$*. In this case, the relation* (7) *holds.*

(iii) W_n *is second-order efficient in the Bahadur sense if and only if* $3J(\theta_0) + 2K(\theta_0) = 0$*. In this case, the relation* (7) *holds for* c_W*.*

2.2. Gaussian stationary processes

Suppose that $\boldsymbol{X}_n = (X_1,\ldots,X_n)'$ is an observed stretch of a Gaussian stationary process with mean 0 and spectral density $f_\theta(\lambda) > 0$, where θ is an unknown parameter. The methodology and the results given here are parallel to the i.i.d. case.

Denote the $n\times n$ Toeplitz matrix associated with $h(\lambda)$ by

$$T_n(h) = \left(\int_{-\pi}^{\pi}\exp\{i(s-t)\lambda\}\,h(\lambda)\,d\lambda\right),$$

$s,t = 1,\ldots,n$, where $h(\lambda)$ is assumed to be an integrable real symmetric function on $[-\pi,\pi]$ (not necessarily nonnegative). The joint density of $\boldsymbol{X}_n = (X_1,\ldots,X_n)'$ from the above Gaussian stationary process is given by $p_n(\boldsymbol{X}_n;\theta) = \exp\{\ell_n(\theta)\}$, where

$$\ell_n(\theta) = -\frac{n}{2}\,\log 2\pi - \frac{1}{2}\,\log\det T_n(f_\theta) - \frac{1}{2}\,\boldsymbol{X}_n'\,T_n(f_\theta)^{-1}\boldsymbol{X}_n \tag{8}$$

is the (exact) log-likelihood of $\boldsymbol{X}_n = (X_1,\ldots,X_n)'$. In this case, Rao's statistic is defined by

$$R_n = n^{-1/2}\frac{\partial}{\partial\theta}\ell_n(\theta)\Big|_{\theta=\theta_0} \quad \text{or} \quad R_n = n^{-1/2}\left|\frac{\partial}{\partial\theta}\ell_n(\theta)\right|_{\theta=\theta_0}$$

and Wald's statistic is

$$W_n = \sqrt{n}(\widehat{\theta}_{n,\mathrm{ML}} - \theta_0) \quad \text{or} \quad W_n = \sqrt{n}|\widehat{\theta}_{n,\mathrm{ML}} - \theta_0|\,,$$

where $\widehat{\theta}_{n,\mathrm{ML}} = \arg\max_\theta \ell_n(\theta)$. Since several authors often use Whittle's (quasi) log-likelihood given by

$$\begin{aligned}\overline{\ell}_n(\theta) &= -\frac{n}{4\pi}\int_{-\pi}^{\pi}\left\{\log f_\theta(\lambda) + \frac{I_n(\lambda)}{f_\theta(\lambda)}\right\} d\lambda \\ &= -\frac{n}{4\pi}\int_{-\pi}^{\pi} \log f_\theta(\lambda)\, d\lambda - \frac{1}{2}\,\boldsymbol{X}_n'\, T_n(f_\theta^{-1}/(4\pi^2))\boldsymbol{X}_n\,,\end{aligned}$$

where

$$I_n(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X_t e^{-it\lambda}\right|^2$$

is the periodogram of $\boldsymbol{X}_n = (X_1, \ldots, X_n)'$, we also examine quasi versions of Rao's statistics qR_n and Wald's statistic qW_n, with $\ell_n(\theta)$ replaced by $\overline{\ell}_n(\theta)$.

Now, we introduce D, a space of functions on $[-\pi, \pi]$ defined by

$$D = \Big\{h(\lambda) = \sum_{u=-\infty}^{\infty} a(u)\exp(-iu\lambda) \,:\, a(u) = a(-u), \sum_{u=-\infty}^{\infty}|a(u)| < \infty\Big\}\,,$$

and set down the following conditions.

(B1) If $\theta \neq \theta'$, then $f_\theta(\lambda) \neq f_{\theta'}(\lambda)$ on a set of positive Lebesgue measure.

(B2) The spectral density $f_\theta(\lambda) \in D$ is bounded away from zero and four times continuously differentiable with respect to θ, and the derivatives $f_\theta^{(i)}(\lambda)$, $i = 1, 2, 3, 4$, belong to D, where $f_\theta^{(i)}(\lambda) = (\partial/\partial\theta)^i f_\theta(\lambda)$.

(B3) The limit of the averaged Fisher information $I^*(\theta)$ is positive for all $\theta \in \Theta$, where

$$I^*(\theta) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\frac{f_\theta^{(1)}(\lambda)}{f_\theta(\lambda)}\right\}^2 d\lambda\,.$$

Define $J^*(\theta), \ldots, H^*(\theta)$ by

$$\begin{aligned}J^*(\theta) &= \frac{1}{4\pi}\int_{-\pi}^{\pi} F_{1,\theta}(\lambda)F_{2,\theta}(\lambda)\,d\lambda\,, & K^*(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} F_{1,\theta}^3(\lambda)\,d\lambda\,,\\ L^*(\theta) &= \frac{1}{4\pi}\int_{-\pi}^{\pi} F_{1,\theta}(\lambda)F_{3,\theta}(\lambda)\,d\lambda\,, & M^*(\theta) &= \frac{1}{4\pi}\int_{-\pi}^{\pi} F_{2,\theta}^2(\lambda)\,d\lambda\,,\\ N^*(\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} F_{1,\theta}^2(\lambda)F_{2,\theta}(\lambda)\,d\lambda\,, & H^*(\theta) &= \frac{3}{2\pi}\int_{-\pi}^{\pi} F_{1,\theta}^4(\lambda)\,d\lambda,\end{aligned}$$

with

$$F_{1,\theta}(\lambda) = \frac{f_\theta^{(1)}(\lambda)}{f_\theta(\lambda)}\,,\ F_{2,\theta}(\lambda) = \frac{f_\theta^{(2)}(\lambda)}{f_\theta(\lambda)} - 2\left\{\frac{f_\theta^{(1)}(\lambda)}{f_\theta(\lambda)}\right\}^2,$$

$$F_{3,\theta}(\lambda) = 6\Big\{\frac{f_\theta^{(1)}(\lambda)}{f_\theta(\lambda)}\Big\}^3 - 6\frac{f_\theta^{(1)}(\lambda)}{f_\theta(\lambda)}\frac{f_\theta^{(2)}(\lambda)}{f_\theta(\lambda)} + \frac{f_\theta^{(3)}(\lambda)}{f_\theta(\lambda)}\,.$$

Theorem 3. *The conclusions in Theorems* 1 *and* 2 *hold for a stationary Gaussian process, if* $I,\ldots,H$ *and* $K(\theta,\theta_0)$ *are replaced by* $I^*,\ldots,H^*$ *and*

$$K^*(\theta,\theta_0) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\Big\{\log\frac{f_{\theta_0}(\lambda)}{f_\theta(\lambda)} - 1 + \frac{f_\theta(\lambda)}{f_{\theta_0}(\lambda)}\Big\}\,d\lambda\,.$$

Remark 1. Although Raghavachari (1970) and Bahadur (1971, p.29) established inequality (2) for the i.i.d. case, their arguments are applicable to the non-i.i.d. case if the log-likelihood ratio $n^{-1}\log\{p_n(\boldsymbol{X}_n;\theta)/p_n(\boldsymbol{X}_n;\theta_0)\}$ converges with P_θ probability one to the limit of the averaged Kullback-Leibler information, where $p_n(\boldsymbol{x};\theta)$ is the joint density of $\boldsymbol{X}_n = (X_1,\ldots,X_n)'$. In the time series case where $\log p_n(\boldsymbol{X}_n;\theta) = \ell_n(\theta)$ is given by (8), we can show that with P_θ probability one,

$$n^{-1}\log\frac{p_n(\boldsymbol{X}_n;\theta)}{p_n(\boldsymbol{X}_n;\theta_0)} \to \lim_{n\to\infty} E_\theta\Big[n^{-1}\log\frac{p_n(\boldsymbol{X}_n;\theta)}{p_n(\boldsymbol{X}_n;\theta_0)}\Big] = K^*(\theta,\theta_0)\,.$$

In fact, this convergence follows from Theorem A.3 in Appendix (see also Hannan (1973, Lemmas 1 and 4)) and the asymptotic expression of log determinant of the Toeplitz matrix by Grenander and Szegö (1984, p.65, (12)):

$$\lim_{n\to\infty}\frac{1}{n}\log\det T_n(h) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\log 2\pi h(\lambda)\,d\lambda \tag{9}$$

if $h(\lambda)$ is a nonnegative integrable function on $[-\pi,\pi]$ and satisfies $\int_{-\pi}^{\pi}\log h(\lambda)\,d\lambda \neq -\infty$. (It should be noted that the Fourier coefficients defined by Grenander and Szegö (1984, p.37, (1)) are divided by 2π.) Then a straightforward extension of Bahadur (1971, p.29) yields that even in a Gaussian stationary process, inequality (2) holds if $K(\theta,\theta_0)$ is replaced by $K^*(\theta,\theta_0)$. Hence, the optimal Bahadur slope in a Gaussian stationary process is $2K^*(\theta,\theta_0)$. Further, Kakizawa (1997) mentioned that the likelihood ratio statistic $LR_n^* = n^{-1}\sup_{\vartheta\in\Theta}[\ell_n(\vartheta) - \ell_n(\theta_0)]$ has the Bahadur slope $2K^*(\theta,\theta_0)$ at each alternative θ (global optimality of LR_n^*) by showing that

$$\liminf_{n\to\infty} LR_n^* \geq K^*(\theta,\theta_0) \text{ with } P_\theta \text{ probability one,}$$

and that, given $\varepsilon > 0$, there exists a positive integer m such that

$$P_{\theta_0}(LR_n^* \geq t) \leq m(1+\varepsilon)^n \mathrm{e}^{-nt}$$

for all $-\infty \leq t \leq \infty$ and all n.

Remark 2. Taniguchi (1996) treats a subject similar to ours. However, he only shows that the weak Bahadur slopes of Rao and Wald statistics (one-sided case) are identical up to order ε^3, and does not discuss the optimal slope in time series. He mentions that the third order term of the Bahadur slope is nothing but $-1/3$ connection coefficient in statistical differential geometry.

3. Proofs of Main Results

Bahadur (1971, p.27) proved that a given test statistic T_n has the slope $c(\theta) = 2g\{b(\theta)\}$ at alternative θ, if there exists an open interval I such that

(C1) $\lim_{n\to\infty} n^{-1} \log P_{\theta_0}(T_n \geq n^{1/2}x) = -g(x)$ for all $x \in I$, where g is a continuous function on I, and

(C2) $\lim_{n\to\infty} n^{-1/2}T_n = b(\theta)$ with probability one, or in probability, when θ obtains, where $b(\theta)$ belongs to I.

Therefore we only have to check that R_n and W_n satisfy these conditions.

We need the following basic lemmas.

Lemma 4. *Under* (A1) *and* (A2), $\sup_t\{xt - \log m(t)\} = x\,\tau(x) - \log m\{\tau(x)\}$ *is* $\mathcal{C}^\infty$ *on* $\{m'(t)/m(t) : t \in (-u,u)\} = \Omega$ *(say), where* $t = \tau(x)$ *is the unique solution of* $m'(t)/m(t) = x$.

Proof. It is easy to see that the second derivative of $\log m(t)$ is positive for $t \in (-u,u)$, which implies that the inverse function Λ^{-1} of $m'(t)/m(t) = \Lambda(t)$ (say) is well-defined and $\tau(x) = \Lambda^{-1}(x)$ for $x \in \Omega$. Further Λ is $\mathcal{C}^\infty$, and so is Λ^{-1}.

Lemma 5. *Under* (A1) *and* (A2),

$$x\,\tau(x) - \log m\{\tau(x)\} = \frac{1}{2\mu_2}\,x^2 - \frac{\mu_3}{6\mu_2^3}\,x^3 + \Big(\frac{\mu_3^2}{8\mu_2^5} - \frac{\mu_4 - 3\mu_2^2}{24\mu_2^4}\Big)x^4 + \mathrm{o}(x^4) \quad (10)$$

as $x \to 0$.

Proof. By the fundamental theorem of implicit functions we can show that there exists a unique single-valued function $t = \tau(x)$ such that $x = m'\{\tau(x)\}/m\{\tau(x)\}$, $\tau(0) = 0$ and $(d/dx)\tau(0) = 1/\mu_2$, which means that $\tau(x) = x/\mu_2 + \mathrm{o}(x)$ as $x \to 0$. Now, let

$$\tau(x) = \frac{x}{\mu_2} + bx^2 + cx^3 + \mathrm{o}(x^3)\,.$$

Since

$$\begin{aligned} m'\{\tau(x)\} &= \mu_2\tau(x) + \frac{\mu_3}{2}\,\tau(x)^2 + \frac{\mu_4}{6}\,\tau(x)^3 + \mathrm{o}(\tau(x)^3) \\ &= x + \Big(b\mu_2 + \frac{\mu_3}{2\mu_2^2}\Big)x^2 + \Big(c\mu_2 + \frac{b\mu_3}{\mu_2} + \frac{\mu_4}{6\mu_2^3}\Big)x^3 + \mathrm{o}(x^3)\,, \end{aligned}$$

$$x\, m\{\tau(x)\} = x\Big[1 + \frac{\mu_2}{2}\tau(x)^2 + \mathrm{o}(\tau(x)^2)\Big]$$
$$= x + \frac{1}{2\mu_2}\, x^3 + \mathrm{o}(x^3)\,,$$

we obtain

$$b = -\frac{\mu_3}{2\mu_2^3}\,,\ c = -\frac{\mu_4 - 3\mu_2^2}{6\mu_2^4} + \frac{\mu_3^2}{2\mu_2^5},$$

and (10) follows immediately from the Taylor expansion.

Proof of Theorem 1. By using the fundamental Chernoff Large Deviation Theorem and the Strong Law of Large Numbers, it is easily checked that R_n satisfies Bahadur's conditions with $I_R = \Omega \bigcap (0,\infty)$, $g_R(x) = x\,\tau(x) - \log m\{\tau(x)\}$ and $b_R(\theta) = E_\theta[\ell^{(1)}(X;\theta_0)]$. Since $E_\theta[\ell^{(1)}(X;\theta_0)] = I(\theta_0)(\theta-\theta_0) + \mathrm{o}(\theta-\theta_0)$, see (11) below, there exists a small $\delta > 0$ such that $E_\theta[\ell^{(1)}(X;\theta_0)]$ belongs to I_R for $\theta \in (\theta_0, \theta_0+\delta)$. Thus, the Bahadur slope of R_n at such an alternative θ exists and is given by $c_R(\theta) = 2g_R\{b_R(\theta)\}$. The expansion (6) follows from (4), (10) and

$$E_\theta[\ell^{(1)}(X;\theta_0)] = E_{\theta_0}[\ell^{(1)}(X;\theta_0)\exp\{\ell(X;\theta_0+\varepsilon) - \ell(X;\theta_0)\}]$$
$$= I\varepsilon + \frac{1}{2}\,(J+K)\varepsilon^2 + \frac{1}{6}\,(L+3N+H)\varepsilon^3 + \mathrm{o}(\varepsilon^3)\,. \qquad (11)$$

Result (ii) is derived from (5) and (6).

The Bahadur slope of W_n is closely related to the exponential rate of the MLE discussed by Fu (1982). He derived the local expansion of $\lim_{n\to\infty} n^{-1}\log P_{\theta_0}(\widehat{\theta}_{n,\mathrm{ML}} > \theta_0 + x) = \log \rho_1(x)$ $(x > 0)$ in terms of Fisher's notations

$$\mu_{ijk} = E_{\theta_0}\Big[\Big(\frac{p^{(1)}}{p}\Big)^i\Big(\frac{p^{(2)}}{p}\Big)^j\Big(\frac{p^{(3)}}{p}\Big)^k\Big],$$

with $p^{(s)} = (\partial/\partial\theta)^s p(X;\theta)|_{\theta=\theta_0}$, $s = 0,1,2,3$. For details on ρ_1, see Lemma 3.2 in Fu (1982). By noting the relations

$$\mu_{110} = J + K\,, \quad \mu_{300} = K\,, \quad \mu_{400} = H + 3I^2$$
$$\mu_{210} = N + H + 2I^2\,, \quad \mu_{101} = L + 3N + H\,,$$
$$\mu_{020} = M + 2N + H + 2I^2\,,$$

Fu's (1982, (3.10)) result is expressed as

$$-2\log\rho_1(x) = Ix^2 + \frac{1}{3}\,(3J+2K)x^3$$
$$+\frac{1}{12}\Big(3\frac{J^2}{I} + 4L + 12N + 3H\Big)x^4 + \mathrm{o}(x^4)\,. \qquad (12)$$

We know that $\log \rho_1(x)$ is continuous in $x \in (0, x_0)$ for some $x_0 > 0$. Further, it is well-known that the MLE is strongly consistent. In the testing case, $-2\log \rho_1(\varepsilon)$ is therefore nothing but the Bahadur slope of W_n at $\theta = \theta_0 + \varepsilon$ for sufficiently small $\varepsilon > 0$, which implies result (iii).

Proof of Theorem 2. Recall the latter part of the proof of Theorem 1 above. Fu (1982) also derived the local expansion of $\lim_{n\to\infty} n^{-1}\log P_{\theta_0}(\widehat{\theta}_{n,\mathrm{ML}} < \theta_0 - x) = \log \rho_2(x)$, $x > 0$, which is similar to (12) with x replaced by $-x$. Then, we can see that Wald's statistic $W_n = \sqrt{n}|\widehat{\theta}_{n,\mathrm{ML}} - \theta_0|$ satisfies Bahadur's conditions with $b(\theta) = |\theta - \theta_0|$ and $g(x) = \min\{-\log \rho_1(x), -\log \rho_1(-x)\}$. This implies that for sufficiently small $\varepsilon \neq 0$, $c_W(\theta_0 + \varepsilon) = 2\min\{-\log \rho_1(|\varepsilon|), -\log \rho_1(-|\varepsilon|)\}$. On the other hand, for Rao's statistic, $b(\theta) = |b_R(\theta)|$ and $g(x) = \min[g_R(x), g_R(-x)]$ for x sufficiently close to 0, $x > 0$, where b_R and g_R are given in the proof of Theorem 1. After some calculation, we obtain

$$\begin{aligned} &c_R(\theta_0+\varepsilon) = 2\min[\ g_R\{|b_R(\theta_0+\varepsilon)|\}, g_R\{-|b_R(\theta_0+\varepsilon)|\}\ ] \\ &= \min\Big[I\varepsilon^2 + \frac{1}{3}(3J+2K)\varepsilon^3 + \frac{1}{12}\Big(3\frac{J^2}{I} + 4L + 12N + 3H\Big)\varepsilon^4 + \mathrm{o}(\varepsilon^4)\,, \\ &\qquad I\varepsilon^2 + \frac{1}{3}(3J+4K)\varepsilon^3 + \frac{1}{12}\Big\{3\frac{(J+2K)^2}{I} + 4L + 12N + 3H\Big\}\varepsilon^4 + \mathrm{o}(\varepsilon^4)\Big]\,. \end{aligned}$$

Proof of Theorem 3. To save space, we consider the one-sided case only. The Bahadur slope of Wald's statistic $W_n = \sqrt{n}(\widehat{\theta}_{n,\mathrm{ML}} - \theta_0)$ or $qW_n = \sqrt{n}(\widehat{\theta}_{n,\mathrm{qML}} - \theta_0)$ follows from Kakizawa (1997) and the strong consistency of $\widehat{\theta}_{n,\mathrm{ML}} = \arg\max_\theta \ell_n(\theta)$ and $\widehat{\theta}_{n,\mathrm{qML}} = \arg\max_\theta \overline{\ell}_n(\theta)$, e.g., Hannan (1973). We turn to the Bahadur slope of Rao's statistic.

Even in the non-i.i.d. case we only have to check conditions (C1) and (C2), since the formulation by Bahadur (1971, p.27) is essentially independent of the i.i.d. assumption. We need, however, a large deviation theorem and almost sure convergence of quadratic forms in Gaussian stationary processes, which will be discussed in Appendix. Set $Z_n = (\partial/\partial\theta)\ell_n(\theta)|_{\theta=\theta_0}$ or $(\partial/\partial\theta)\overline{\ell}_n(\theta)|_{\theta=\theta_0}$. It follows from Theorems A.1 and A.2 (i) in Appendix that there exists an $r > 0$ such that for all $t \in [-r, r]$, $\lim_{n\to\infty} n^{-1}\log E_{\theta_0}\{\exp(tZ_n)\} = \phi(t)$, where

$$\phi(t) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\log\Big\{1 - tf_{\theta_0}(\lambda)^{-1}\frac{\partial}{\partial\theta}f_{\theta_0}(\lambda)\Big\}\,d\lambda - \frac{t}{4\pi}\int_{-\pi}^{\pi} f_{\theta_0}(\lambda)^{-1}\frac{\partial}{\partial\theta}f_{\theta_0}(\lambda)\,d\lambda\,. \tag{13}$$

By (B3), $\{\lambda \,|\, (\partial/\partial\theta)f_{\theta_0}(\lambda) \neq 0\}$ has positive Lebesgue measure, hence $\phi(t)$ is strictly convex. This means that for $x \in (\ 0, (d/dt)\phi(t)|_{t=r}\ ) = I_R^*$ (say), there exists a unique $0 < t(x) < r$ such that

$$\frac{d}{dt}\phi(t)\Big|_{t=t(x)} = x$$

and

$$\lim_{n\to\infty}\frac{1}{n}\log P_{\theta_0}(Z_n \ge nx) = \phi\{t(x)\} - x\,t(x),$$

see Theorem A.2 (ii) in Appendix. Therefore, both R_n and qR_n satisfy condition (C1) with $g_R^*(x) = x\,t(x) - \phi\{t(x)\}$. As in Lemma 5, this function has a four-term expansion at $x = 0$ given by

$$x\,t(x) - \phi\{t(x)\} = \frac{1}{2I^*}\,x^2 - \frac{K^*}{6(I^*)^3}\,x^3 + \left\{\frac{(K^*)^2}{8(I^*)^5} - \frac{H^*}{24(I^*)^4}\right\}x^4 + \mathrm{o}(x^4)\,.$$

For (C2), use Theorems A.1 and A.3 in the Appendix:

$$\begin{aligned}\lim_{n\to\infty} n^{-1/2}R_n &= \lim_{n\to\infty} n^{-1/2}qR_n \\ &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[f_\theta(\lambda)\,f_{\theta_0}(\lambda)^{-2}\frac{\partial}{\partial\theta}f_{\theta_0}(\lambda) - f_{\theta_0}(\lambda)^{-1}\frac{\partial}{\partial\theta}f_{\theta_0}(\lambda)\right]d\lambda \\ &= b_R^*(\theta) \quad \text{(say)} \end{aligned} \tag{14}$$

with P_θ probability one. Let $\theta = \theta_0 + \varepsilon$ for sufficiently small $\varepsilon > 0$. Then it is easy to see

$$b_R^*(\theta_0+\varepsilon) = I^*\varepsilon + \frac{1}{2}(J^* + K^*)\varepsilon^2 + \frac{1}{6}(L^* + 3N^* + H^*)\varepsilon^3 + \mathrm{o}(\varepsilon^3)\,,$$

which has a similar structure as (11) in the proof of Theorem 1. To get the four-term expansion of $K^*(\theta_0+\varepsilon,\theta_0)$, it is useful to note that

$$\frac{f_\theta(\lambda)}{f_{\theta+\varepsilon}(\lambda)} = 1 - F_{1,\theta}(\lambda)\varepsilon - \frac{1}{2}F_{2,\theta}(\lambda)\varepsilon^2 - \frac{1}{6}F_{3,\theta}(\lambda)\varepsilon^3 + \mathrm{O}(\varepsilon^4)\,.$$

4. Case Studies

The results in Section 2 are true for general models, including Gaussian stationary processes. To illustrate these results, we present two examples and consider the one-sided case only.

Example 1. Let $(X_i, Y_i)'$, $i = 1, \ldots, n$, be a sequence of i.i.d. random vectors with common density function

$$p(x, y; \beta) = \exp\Big(-\beta x - \frac{1}{\beta}\,y\Big)\,, \quad x, y > 0\,.$$

This is a curved exponential family with Fisher information $I(\beta) = 2/\beta^2$ and statistical curvature $\gamma(\beta) = 1/\sqrt{2}$. We wish to test $H : \beta = \beta_0$ against $A_1 : \beta >$

β_0, where $\beta_0 > 0$ is given. In this case, the Bahadur slopes of Rao and Wald statistics at alternative β are obtained explicitly as follows:

$$c_R(\beta) = -2 + 2\Big\{1 + u^2\Big(\frac{2+u}{1+u}\Big)^2\Big\}^{1/2} - 2\log\frac{1}{2}\Big[1 + \Big\{1 + u^2\Big(\frac{2+u}{1+u}\Big)^2\Big\}^{1/2}\Big], \tag{15}$$

$$c_W(\beta) = 4\log\Big(1 + u + \frac{u^2}{2}\Big) - 4\log(1+u), \tag{16}$$

where $u = (\beta - \beta_0)/\beta_0$. Since the Kullback-Leibler information in this model is $K(\beta, \beta_0) = u^2/(1+u)$, it is easy to see that

$$\begin{aligned}\lim_{\varepsilon\to 0}\varepsilon^{-i}\,[2K(\beta_0+\varepsilon,\beta_0) - c_R(\beta_0+\varepsilon)] &= \lim_{\varepsilon\to 0}\varepsilon^{-i}\,[2K(\beta_0+\varepsilon,\beta_0) - c_W(\beta_0+\varepsilon)]\\ &= \begin{cases}0\,, & i = 1,2,3\\ 1/(2\beta_0^4)\,, & i = 4\,,\end{cases}\end{aligned}$$

and the RHS for $i = 4$ is equal to $(1/4)\{I(\beta_0)\gamma(\beta_0)\}^2$, which agrees with Theorem 1. Interestingly, we find that Rao's statistic is better than Wald's statistic in the sense that

$$\lim_{\varepsilon\to 0}\varepsilon^{-i}\,[c_R(\beta_0+\varepsilon) - c_W(\beta_0+\varepsilon)] = \begin{cases}0\,, & i = 1,2,3,4,5\\ 2/(3\beta_0^6)\,, & i = 6\,.\end{cases}$$

This provides an example in which the Bahadur slope of Rao's statistic coincides with that of Wald's statistic up to order ε^5.

Proof of (15). It is easy to see that the MGF of $(\partial/\partial\beta)\log p(X,Y;\beta)|_{\beta=\beta_0}$ under $\beta = \beta_0$ is given by

$$m(t) = \frac{\beta_0^2}{\beta_0^2 - t^2}\,.$$

For any $z > 0$, the equation $(d/dt)\log m(\widehat{t}) = z$ has a unique solution

$$\widehat{t} = \frac{1}{z}\,\{-1 + (1 + z^2\beta_0^2)^{1/2}\}\,.$$

It follows from Chernoff's Large Deviation Theorem that

$$\lim_{n\to\infty} n^{-1}P_{\beta_0}(R_n \geq n^{1/2}z) = \log m(\widehat{t}) - \widehat{t}z\,.$$

On the other hand, the Strong Law of Large Numbers implies

$$n^{-1/2}R_n \xrightarrow{a.s.} \frac{\beta^2 - \beta_0^2}{\beta\beta_0^2} > 0$$

under alternative $\beta > \beta_0$. Thus, (C1) and (C2) are checked.

Proof of (16). It is easy to see that the MLE for β is given by $\widehat{\beta}_{n,\mathrm{ML}} = (\overline{Y}/\overline{X})^{1/2}$ and for $z > 0$,

$$\lim_{n\to\infty} n^{-1} P_{\beta_0}(\widehat{\beta}_{n,\mathrm{ML}} - \beta_0 \geq z) = \lim_{n\to\infty} n^{-1} P_{\beta_0}\{\overline{Y} - (\beta_0 + z)^2 \overline{X} \geq 0\}\,. \tag{17}$$

The MGF of $Y - (\beta_0 + z)^2 X$ under $\beta = \beta_0$ is

$$m(t) = \frac{\beta_0}{\beta_0 + (\beta_0 + z)^2 t} \frac{1}{1 - \beta_0 t},$$

and the equation $(d/dt) \log m(\widehat{t}) = 0$ has a unique solution

$$\widehat{t} = \frac{(\beta_0 + z)^2 - \beta_0^2}{2\beta_0(\beta_0 + z)^2}$$

for any $z > 0$. Thus, Chernoff's Large Deviation Theorem implies that the RHS of (17) is equal to

$$\log m(\widehat{t}) = -2 \log\Big(1 + \frac{z}{\beta_0} + \frac{z^2}{2\beta_0^2}\Big) + 2 \log\Big(1 + \frac{z}{\beta_0}\Big)\,.$$

Further, it is easily checked that $\widehat{\beta}_{n,\mathrm{ML}}$ converges with P_β probability one to β.

Example 2. Let $\boldsymbol{X}_n = (X_1, \ldots, X_n)'$ be an observed stretch from a Gaussian AR(1) process $X_t = \theta X_{t-1} + U_t$, $t = \ldots, -1, 0, 1, \ldots$, where U_t is a sequence of independent $N(0, 1)$ random variables and $|\theta| < 1$. Consider the problem of testing $H : \theta = \theta_0$ against $A_1 : \theta > \theta_0$, where θ_0 is given. In the context of estimation, Kakizawa (1998) showed that the exponential rates of upper tail probabilities of several estimators $\widehat{\theta}_n$ (e.g. MLE, least squares estimator, Daniels's estimator, the Yule-Walker estimator, ...) are identical:

$$\lim_{n\to\infty} n^{-1} \log P_{\theta_0}(\widehat{\theta}_n \geq \theta_0 + \varepsilon) = \frac{1}{2} \log \psi(\theta_0 + \varepsilon) \quad \text{with } \psi(r) = \frac{1 - r^2}{1 - 2\theta_0 r + \theta_0^2}\,,$$

when $\varepsilon > 0$ is sufficiently small. In the context of testing, this means that several Wald type tests based on the estimators $\widehat{\theta}_n$ have Bahadur slope $c_W(\theta) = -\log \psi(\theta)$ at an alternative θ close to θ_0, $\theta > \theta_0$. It is easy to see that the limit of the averaged Kullback-Leibler information in a Gaussian AR(1) model is $K^*(\theta, \theta_0) = (\theta - \theta_0)^2/\{2(1 - \theta^2)\}$, see Kakizawa (1998). Hence

$$\lim_{\varepsilon\to 0} \varepsilon^{-i} \left[2K^*(\theta_0 + \varepsilon, \theta_0) - c_W(\theta_0 + \varepsilon)\right] = \begin{cases} 0\,, & i = 1, 2, 3 \\ (1/4)\{I^*(\theta_0)\gamma^*(\theta_0)\}^2\,, & i = 4\,, \end{cases}$$

where $I^*(\theta) = 1/(1 - \theta^2)$ is the limit of the averaged Fisher information and $\gamma^*(\theta) = \sqrt{2}$ is a counterpart of the statistical curvature in a Gaussian AR(1) model, e.g. Taniguchi (1991, p.95). For Rao's statistic, by setting

$$f_\theta(\lambda) = \frac{1}{2\pi}\,\frac{1}{|1-\theta e^{i\lambda}|^2}\,,$$

(13) and (14) in the proof of Theorem 3 become

$$b(\theta) = \frac{\theta-\theta_0}{1-\theta^2}$$

and

$$\phi(t) = -\frac{1}{2}\,\log\frac{1}{2}\left[\alpha(t) + \{\alpha^2(t) - 4(\theta_0+t)^2\}^{1/2}\right],$$

where $\alpha(t) = 1+\theta_0^2+2\theta_0 t$. Therefore, the Bahadur slope of Rao's statistic is implicitly given by

$$c_R(\theta) = 2b(\theta)\,t_b + \log\frac{1}{2}\left[\alpha(t_b) + \{\alpha^2(t_b) - 4(\theta_0+t_b)^2\}^{1/2}\right],$$

where t_b is the unique solution of $(d/dt)\phi(t)|_{t=t_b} = b(\theta)$. After tedious calculation, we can show that

$$\lim_{\varepsilon\to 0}\varepsilon^{-i}\,[c_R(\theta_0+\varepsilon) - c_W(\theta_0+\varepsilon)] = \begin{cases} 0\,, & i=1,2,3,4 \\ 2\theta_0/(1-\theta_0^2)^3\,, & i=5\,. \end{cases}$$

This indicates that Rao's statistic is better than Wald's statistic if $\theta_0 > 0$ and vice versa. That is, uniform superiority of the ε^5-term in the Bahadur slope does not hold generally in the comparison of Rao and Wald statistics for the one-sided case. It should be remarked that in the two-sided case, this phenomenon occurs in the ε^3-term, see Theorem 2.

Appendix

In this Appendix we present a large deviation theorem and almost sure convergence of quadratic forms in Gaussian stationary processes. These results have independent interest.

First we have a result about the trace of Toeplitz matrices.

Theorem A.1. *Assume that $f_j(\lambda) \in D$, $j=1,\ldots,s$, are strictly positive on $[-\pi,\pi]$, and that $g_j(\lambda)\in D$, $j=1,\ldots,s$. Then*

$$\lim_{n\to\infty}\frac{1}{n}\,\mathrm{tr}\Big\{T_n(f_1)^{-1}T_n(g_1)T_n(f_2)^{-1}T_n(g_2)\cdots T_n(f_s)^{-1}T_n(g_s)\Big\}$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{g_1(\lambda)\cdots g_s(\lambda)}{f_1(\lambda)\cdots f_s(\lambda)}\,d\lambda\,.$$

Proof. Under stronger conditions, Coursol and Dacunha-Castelle (1982) and Taniguchi (1991, p.16) derived a similar formula with rate $\mathrm{O}(n^{-1})$. The asymptotics we consider here do not need the evaluation of the remainder term, so we only assume the summability condition on the Fourier coefficients $a(u)$ for $h(\lambda) \in D$.

Define the Euclidean norm and spectral norm of a matrix A by $||A||_E = [\mathrm{tr}(A^*A)]^{1/2}$ and $||A|| = \sup_{|z|=1}[\boldsymbol{z}^*A^*A\boldsymbol{z}]^{1/2}$, respectively, where * denotes conjugate transpose. Our proof is mainly based on properties of asymptotically equivalent matrices (e.g. Graybill (1983, Definition 5.6.5)). Along the line of Davies (1973), it is easy to see that for all $h(\lambda) \in D$, Toeplitz matrices $T_n(h)$ commute asymptotically in the sense that

$$\lim_{n\to\infty} n^{-\frac{1}{2}}||U_nT_n(h)U_n^* - 2\pi D_n(h)||_E = 0\,, \tag{A.1}$$

where $D_n(h)$ is the $n \times n$ diagonal matrix with $h(\lambda_k)$ as the kth diagonal element (we define $\lambda_k = 2\pi k/n$), and U_n is the $n \times n$ Fourier unitary matrix with $n^{-\frac{1}{2}}\mathrm{e}^{i2\pi st/n}$ as the (s,t)th element. Further, we can prove that the spectral norm $||T_n(h)||$ of $T_n(h)$ is bounded by $\max_\lambda |2\pi h(\lambda)|$. Thus, from (A.1), $U_nT_n(h)U_n^*$ is asymptotically equivalent to $2\pi D_n(h)$. Since $f(\lambda) > 0$ implies the nonsingularity of $T_n(f)$, and $||T_n(f)^{-1}|| \le 1/\min_\lambda\{2\pi f(\lambda)\}$, e.g. Brockwell and Davis (1991, Proposition 4.5.3), it follows from Theorem 5.6.11 (2) in Graybill (1983) that $\{U_nT_n(f)U_n^*\}^{-1}$ is asymptotically equivalent to $\{2\pi D_n(f)\}^{-1}$. Therefore, from Theorem 5.6.12 (1) in Graybill (1983), we can show that $U_nT_n(f_1)^{-1}T_n(g_1)\cdots T_n(f_s)^{-1}T_n(g_s)U_n^*$ is asymptotically equivalent to $D_n(f_1)^{-1}D_n(g_1)\cdots D_n(f_s)^{-1} D_n(g_s) = D_n$ (say). Noting that

$$\frac{1}{n}\,\mathrm{tr}(D_n) = \frac{1}{n}\sum_{k=1}^{n}\frac{g_1(\lambda_k)\cdots g_s(\lambda_k)}{f_1(\lambda_k)\cdots f_s(\lambda_k)} \to \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{g_1(\lambda)\cdots g_s(\lambda)}{f_1(\lambda)\cdots f_s(\lambda)}\,d\lambda\,,$$

the result is a consequence of Theorem 2.1 in Gray (1972). Note that (9) in subsection 2.2 is also shown by the above approach by using asymptotic equivalent matrices.

Using Theorem A.1, we can evaluate the limits of the averaged cumulant generating functions of quadratic forms in Gaussian stationary processes, which plays an important role in the large deviation theorem.

Theorem A.2. *Let $\boldsymbol{X}_n = (X_1,\ldots,X_n)'$ be an observed stretch of a zero mean Gaussian stationary process with spectral density $f(\lambda) \in D$, strictly positive. Define the generalized quadratic form*

$$Z_n = \frac{1}{2}\Big\{\boldsymbol{X}_n'\,T_n(g_1)^{-1}\,T_n(g_2)\,T_n(g_1)^{-1}\boldsymbol{X}_n + c_n\Big\}\,,$$

where $g_1(\lambda) \in D$ *is strictly positive,* $g_2(\lambda) \in D$*, and* $\{c_n\}$ *is a sequence such that*

$$\lim_{n\to\infty} \frac{c_n}{n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} g_3(\lambda)\, d\lambda$$

with an integrable real function $g_3(\lambda)$*. Then,* (i) *there exists* $r > 0$ *such that for all* $t \in [-r, r]$, $\lim_{n\to\infty} n^{-1} \log E\{\exp(tZ_n)\} = \phi_Z(t)$*, where*

$$\phi_Z(t) = -\frac{1}{4\pi} \int_{-\pi}^{\pi} \log\left\{1 - t\frac{f(\lambda)\, g_2(\lambda)}{g_1^2(\lambda)}\right\} d\lambda + \frac{t}{4\pi} \int_{-\pi}^{\pi} g_3(\lambda)\, d\lambda\,.$$

In addition, (ii) *if* $g_3(\lambda) = -f(\lambda)g_2(\lambda)/g_1^2(\lambda)$ *and* $\{\lambda \,|\, g_2(\lambda) \neq 0\}$ *has positive Lebesgue measure, then*

$$\lim_{n\to\infty} \frac{1}{n} \log P(Z_n \geq nx) = \phi_Z\{t(x)\} - x\, t(x) < 0 \qquad \textit{for } 0 < x < \frac{d}{dt}\phi_Z(t)\Big|_{t=r}, \tag{A.2a}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log P(Z_n \leq nx) = \phi_Z\{t(x)\} - x\, t(x) < 0 \qquad \textit{for } \frac{d}{dt}\phi_Z(t)\Big|_{t=-r} < x < 0, \tag{A.2b}$$

where $t(x)$ *is the unique solution of*

$$\frac{d}{dt}\, \phi_Z(t)\Big|_{t=t(x)} = x\,. \tag{A.3}$$

Proof. It is easy to see that $\phi_Z(t)$, $|t| \leq r$, is strictly convex and there exists a unique solution $t(x)$ of (A.3) if x belongs to the range of $(d/dt)\phi_Z(t)$, $|t| \leq r$. Note that $t(x)$ is continuous and strictly increasing in x. Since (A.2) is a consequence of a version of a large deviation theorem (e.g. Ihara (1993, p.113)), it suffices to prove (i).

Let

$$A_n = T_n(f)^{1/2}\, T_n(g_1)^{-1}\, T_n(g_2)\, T_n(g_1)^{-1}\, T_n(f)^{1/2}\,.$$

If $|t| < 1/||A_n||$, then $E\{\exp(tZ_n)\}$ exists and

$$\phi_n(t) = \frac{1}{n} \log E\{\exp(tZ_n)\} = \frac{t}{2n}\, c_n - \frac{1}{2n} \log\det(I - tA_n)\,.$$

Now, let $g_2(\lambda) = g_2^+(\lambda) - g_2^-(\lambda)$, where $g_2^+(\lambda)$ and $g_2^-(\lambda)$ are nonnegative. For $h(\lambda) = f(\lambda), g_2^+(\lambda), g_2^-(\lambda)$, we have

$$\begin{aligned} ||T_n(g_1)^{-1/2} T_n(h)^{1/2}||^2 &= ||T_n(h)^{1/2} T_n(g_1)^{-1/2}||^2 \\ &= \sup_{|z|=1} \frac{\boldsymbol{z}'\, T_n(h)\boldsymbol{z}}{\boldsymbol{z}'\, T_n(g_1)\boldsymbol{z}} \end{aligned}$$

$$= \sup_{|z|=1} \frac{\int_{-\pi}^{\pi} h(\lambda) \,|\sum_{j=1}^{n} z_j \mathrm{e}^{-ij\lambda}|^2 \, d\lambda}{\int_{-\pi}^{\pi} g_1(\lambda) \,|\sum_{j=1}^{n} z_j \mathrm{e}^{-ij\lambda}|^2 \, d\lambda}$$
$$\le \frac{\max_\lambda h(\lambda)}{\min_\lambda g_1(\lambda)},$$

hence

$$||A_n|| \le ||T_n(f)^{1/2} T_n(g_1)^{-1/2}||^2 (||T_n(g_2^+)^{1/2} T_n(g_1)^{-1/2}||^2 + ||T_n(g_2^-)^{1/2} T_n(g_1)^{-1/2}||^2)$$
$$\le 1/R \quad \text{(say)}.$$

For all $|t| \le r$ $(0 < r < R)$, an expansion of $\phi_n(t)$ is

$$\frac{t}{2n} c_n + \frac{1}{2n} \sum_{j=1}^{\infty} \frac{t^j}{j} \operatorname{tr}\Big\{T_n(f) T_n(g_1)^{-1} T_n(g_2) T_n(g_1)^{-1}\Big\}^j .$$

From Theorem A.1 we obtain

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{tr}\Big\{T_n(f) T_n(g_1)^{-1} T_n(g_2) T_n(g_1)^{-1}\Big\}^j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Big\{\frac{f(\lambda)\, g_2(\lambda)}{g_1^2(\lambda)}\Big\}^j d\lambda$$

for all positive integers j. Since

$$\sum_{j=1}^{\infty} \frac{t^j}{j} \int_{-\pi}^{\pi} \Big\{\frac{f(\lambda)\, g_2(\lambda)}{g_1^2(\lambda)}\Big\}^j d\lambda = -\int_{-\pi}^{\pi} \log\Big\{1 - t \frac{f(\lambda)\, g_2(\lambda)}{g_1^2(\lambda)}\Big\} d\lambda$$

for $|t| \le r$, we get the result (i).

We conclude this appendix by proving almost sure convergence of the quadratic form $Q_n = \frac{1}{2} \boldsymbol{X}_n' T_n(g_1)^{-1} T_n(g_2) T_n(g_1)^{-1} \boldsymbol{X}_n$.

Theorem A.3. *Under the conditions of Theorem* A.2*,*

$$\lim_{n\to\infty} n^{-1} Q_n = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{f(\lambda)\, g_2(\lambda)}{g_1^2(\lambda)} \, d\lambda$$

with probability one.

Proof. The result holds trivially if $g_2(\lambda) = 0$, so we assume $\{\lambda \,|\, g_2(\lambda) \ne 0\}$ has positive Lebesgue measure. Since Theorem A.1 implies

$$\lim_{n\to\infty} \frac{1}{n} E Q_n = \lim_{n\to\infty} \frac{1}{2n} \operatorname{tr}\Big\{T_n(f) T_n(g_1)^{-1} T_n(g_2) T_n(g_1)^{-1}\Big\} = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{f(\lambda)\, g_2(\lambda)}{g_1^2(\lambda)} d\lambda,$$

it suffices to prove $n^{-1} Z_n \xrightarrow{a.s.} 0$, where $Z_n = Q_n - E Q_n$. From Theorem A.2 (ii), it follows that for all sufficiently small $\varepsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log P(Z_n \ge n\varepsilon) = \phi_Z\{t(\varepsilon)\} - \varepsilon t(\varepsilon) < 0$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log P(Z_n \le -n\varepsilon) = \phi_Z\{t(-\varepsilon)\} - (-\varepsilon)t(-\varepsilon) < 0\,.$$

Hence, there exists a number $N = N(\varepsilon) > 0$ such that $P(|n^{-1}Z_n| \ge \varepsilon) \le \mathrm{e}^{-nN}$ for all sufficiently large n. It then follows from the Borel-Cantelli Lemma that $n^{-1}Z_n \xrightarrow{a.s.} 0$.

References

Akahira, M. and Takeuchi, K. (1981). *Asymptotic Efficiency of Statistical Estimators; Concepts and Higher Order Asymptotic Efficiency.* Lecture Notes in Statistics, Vol. 7. Springer-Verlag, Berlin.

Akritas, M. G. and Kourouklis, S. (1988). Local Bahadur efficiency of score tests. *J. Statist. Plann. Inference* **19**, 187-199.

Bahadur, R. R. (1960). Stochastic comparison of tests. *Ann. Math. Statist.* **31**, 276-295.

Bahadur, R. R. (1971). *Some Limit Theorems in Statistics.* SIAM, Philadelphia.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* 2nd edition. Springer-Verlag, New York.

Coursol, J. and Dacunha-Castelle, D. (1982). Remarks on the approximation on the likelihood function of a stationary Gaussian process. *Theory Probab. Appl.* **27**, 162-167.

Davies, R. B. (1973). Asymptotic inference in stationary Gaussian time-series. *Adv. Appl. Probab.* **5**, 469-497.

Fu, J. C. (1982). Large sample point estimation: a large deviation theory approach. *Ann. Statist.* **10**, 762-771.

Gray, R. M. (1972). On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE Trans. Inf. Theory* **18**, 725-730.

Graybill, F. A. (1983). *Matrices with Applications in Statistics.* 2nd edition. Wadsworth, California.

Grenander, U. and Szegö, G. (1984). *Toeplitz Forms and Their Applications.* 2nd edition. Chelsea, New York.

Hannan, E. J. (1973). The asymptotic theory of linear time series models. *J. Appl. Probab.* **10**, 130-145.

He, X. and Shao, Q. M. (1996). Bahadur efficiency and robustness of studentized score tests. *Ann. Inst. Statist. Math.* **48**, 295-314.

Ihara, S. (1993). *Information Theory for Continuous Systems.* World Scientific, Singapore.

Kakizawa, Y. (1997). On Bahadur asymptotic efficiency of the maximum likelihood and quasi-maximum likelihood estimators for Gaussian stationary processes. *Discussion Paper Series* (A), No.41. Hokkaido University. To appear *Stochast. Process Appl.* (1999).

Kakizawa, Y. (1998). On exponential rates of estimators of the parameter in the first-order autoregressive process. *Statist. Probab. Lett.* **38**, 355-362.

Mukerjee, R. (1993). Rao's score test: recent asymptotic results. In *Handbook of Statistics*, Vol.11. (Edited by Maddala, Rao and Vinod), 363-379. North-Halland, Amsterdam.

Raghavachari, M. (1970). On a theorem of Bahadur of the rate of convergence of test statistics. *Ann. Math. Statist.* **41**, 1695-1699.

Rao, C. R. and Mukerjee, R. (1997). Comparison of LR, score, and Wald tests in a non-iid setting. *J. Multivariate Anal.* **60**, 99-110.

Sato, T., Kakizawa, Y. and Taniguchi, M. (1998). Large deviation results for statistics of short- and long-memory Gaussian processes. *Austral. & New Zealand J. Statist.* **40**, 17-29.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Taniguchi, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis*. Lecture Notes in Statistics, Vol. 68. Springer-Verlag, Heidelberg.

Taniguchi, M. (1996). On large deviation asymptotics of some tests in time series. *Preprints de l'Institut de Statistique Série Probabilités et Statistique*, No. 52. Universite Libre de Bruxelles. To appear in *J. Statist. Plann. Inference* (1999).

Faculty of Economics, Hokkaido University, Nishi 7, Kita 9, Kita-ku, Sapporo 060–0809, Japan.

E-mail: kakizawa@econ.hokudai.ac.jp

(Received April 1998; accepted March 1999)