Journal of the Korean Statistical Society (1998), **27**: 3, pp 251–264

On Bahadur Efficiency and Bartlett Adjustability of Quasi-LRT Statistics†

Kwan Jeh Lee[1]

ABSTRACT

When the LRT is not feasible, we define quasi-LRT(QLRT) as a modification of the LRT. Under some appropriate conditions the QLRT shares Bahadur optimality and Bartlett Adjustability with the LRT. When we can find maximum likelihood estimator under the null parameter space but not under the unrestricted parameter space, our QLRT is Bahadur optimal as is the LRT. We suggest the stopping rule of the Newton-Raphson iterations for constructing the QLRT statistics which are Bartlett adjustable.

Keywords: Bahadur efficiency; Bartlett adjustability; Asymptotic relative efficiency; Likelihood ratio; Quasi-likelihood ratio

1. Introduction

The most widely known and used general method of testing a hypothesis concerning parameters in a statistical model is the likelihood ratio test(LRT). There are several reasons for the desirability of the LRT such as its asymptotic distribution, Bahadur optimality, and Bartlett adjustability. A problem in carrying out the LRT arises when maximum likelihood estimates are difficult to obtain. Often times the maximization of the likelihood, either under the null and/or under the unrestricted model, is not feasible. In such cases, there are several methods to get estimators that share properties of maximum likelihood estimators(MLEs) in the sense that they are consistent and asymptotically efficient. In testing hypothesis, however, it is known that Wald test and Score test, as competitors of the LRT, does not share Bahadur optimality and Bartlett adjustability, even though their asymptotic distributions are the same as that of the LRT. This creates a need for modification of the LRT in testing hypothesis. There are a series of papers on

†This paper was supported in part by NON-DIRECTED RESEARCH FUND, Korea Research Foundation.

[1]Department of Statistics, Dongguk University, Seoul 100-715, Korea.

Bahadur optimality of the LRT such as Bahadur(1960a, 1960b,1965,1967,1971), and Bahadur and Raghavachari(1970). Barndorff-Nielsen and Cox(1984), and Barndorff-Nielsen and Hall(1988) showed the high-order asymptotic distribution of the LRT statistic.

In the present study we propose a quasi-likelihood ratio test(defined later in this paper) as a modification of the LRT. In some situations the quasi-likelihood ratio test(QLRT) shares Bahadur optimality and Bartlett adjustability with the LRT. In section 2 we define the QLRT and in section 3 and 4 show that some of QLRTs have the Bahadur slop of the LRT, which is called the first-order Bahadur efficient. In section 5 we find the sufficient conditions for Bartlett adjustability of QLRT.

2. Quasi-LRT

Let $X_1, X_2, \cdots$ and X be a sequence of independent, identically distributed random variables. And let X have a density function $f(x,\theta)$ where $\theta \in \Theta$, which is a subset of Euclidean space. The space Θ is called the parameter space. The likelihood function, given a sample with n observations, is denoted by

$$L(\theta|\underline{x}) = \prod_{i=1}^{n} f(x_i, \theta),$$

where $\underline{x} = (x_1, \cdots, x_n)$, and the corresponding log likelihood function is written as

$$l_n(\theta) \equiv \log L(\theta|\underline{x}).$$

The hypotheses on the parameters considered in the study are as follows:

$$H_o : \theta \in \Theta_o \text{ vs. } H_1 : \theta \notin \Theta_o,$$

where the null parameter space Θ_o is a hyperplane of the whole parameter space Θ.

The LRT statistic $\hat{\Lambda}_n$ is defined by

$$\begin{aligned}\hat{\Lambda}_n &= \frac{\sup_{\Theta_o} L(\theta|\underline{x})}{\sup_{\Theta} L(\theta|\underline{x})} \\ &= \frac{L(\hat{\theta}_o|\underline{x})}{L(\hat{\theta}|\underline{x})},\end{aligned}$$

where $\hat{\theta}_o$ and $\hat{\theta}$ are MLEs under the null and the whole (unrestricted) parameter spaces, respectively.

We define quasi-likelihood ratio statistics as follows.

Definition 2.1. *Let Λ_n^* be quasi-likelihood ratio test(QLRT) statistics if*

$$\Lambda_n^* = \frac{L(\theta_o^*|\underline{x})}{L(\theta^*|\underline{x})}$$

where θ_o^* and θ^* are consistent estimators of θ under the null and under the whole parameter spaces, respectively.

A classical result due to Wilks(1938) on the distribution of the LRT statistic $\hat{\Lambda}_n$ is the following. Under suitable regularity conditions, if the hypothesis that a parameter θ lies on an r-dimensional hyperplane of k-dimensional whole parameter space is true, the distribution of $-2\log\hat{\Lambda}_n$ is asymptotically that of χ^2 variate with $k-r$ degrees of freedom.

Under some regularity conditions our QLRT statistics behave like the LRT statistic, in the sense that they converge in law to a χ^2 distribution with the same degree of freedom of the LRT statistic. In fact without additional conditions,

$$-2\log\Lambda_n^* = -2\log\hat{\Lambda}_n + o_p(1),$$

if we use in constructing the QLRT statistic a consistent estimator of the parameter such as $\hat{\theta}-\theta^* = o_p(n^{-1/2})$. When we cannot implement the LRT, the QLRT might be a competitor of Wald test and score test.

3. Bahadur Efficiency

Let $s = (x_1, x_2, \cdots,)$ be a sequence of *iid* observations on X. Let P_θ be the probability distribution of s in its sample space when θ is true. For each $n = 1, 2, \cdots$, let $T_n(s)$ be an extended real-valued measurable function which depends on s only through $(x_1, x_2, \cdots, x_n)$. T_n is to be thought of as a test statistic.

Consider now the testing problem $H_0 : \theta \in \Theta_o$ vs. $H_1 : \theta \in \Theta - \Theta_o$, for which large values of T_n are significant. For each θ, let

$$F_n(t,\theta) = P_\theta(T_n < t), \tag{3.1}$$

and let

$$G_n(t) = \inf\{F_n(t,\theta) : \theta \in \Theta_o\}, \tag{3.2}$$

for $-\infty < t < \infty$. Define

$$H(T_n) = 1 - G_n(T_n(s)). \tag{3.3}$$

$H(T_n)$ has been called the "level attained" by T_n. Thus the "level attained" is a random variable representing the degree to which the test statistic T_n tends to reject H_0. The lower the value of the level attained, the greater the evidence against the null hypothesis H_0. See Bahadur(1967,1971) and Raghavachari(1970) for the behavior of the level attained H. We say that a test sequence $\{T_n\}$ has the exact Bahadur slope $c(\theta)$ when θ is true if

$$\lim_{n\to\infty} \frac{1}{n} \log H(T_n) = -\frac{1}{2}c(\theta) \tag{3.4}$$

with probability one.

In the non-null case the limit $c(\theta)$ may be regarded as a measure of the performance of T_n; the larger the value of $c(\theta)$ the "faster" T_n tends to reject H_0.

Let $\kappa(\theta,\theta_o)$ be the Kullback-Leibler information number, that is,

$$\kappa(\theta,\theta_o) = E_o(log\frac{f(X,\theta)}{f(X,\theta_o)}).$$

For each $\theta \in \Theta$, let

$$J(\theta) = \inf\{\kappa(\theta,\theta_o) : \theta_o \in \Theta_o\}.$$

Then $J(\theta)$ is well defined over Θ, and $0 \le J \le \infty$ and $J = 0$ on Θ_o. Bahadur(1971) showed that for each θ in $\Theta - \Theta_o$,

$$\liminf_{n\to\infty} \frac{1}{n} \log H(T_n) \ge -J(\theta), \tag{3.5}$$

when θ obtains.

Since small values of the LRT statistic $\hat{\Lambda}_n$ are significant, we consider instead an equivalent statistic, $\hat{T}_n$ say, where $\hat{T}_n$ is a strictly decreasing function of $\hat{\Lambda}_n$ for each n. Therefore we choose

$$\hat{T}_n = -\frac{1}{n}\log \Lambda_n.$$

Let $\hat{F}_n$, $\hat{G}_n$, and $H(\hat{T}_n)$ as in (3.1), (3.2), and (3.3), respectively. That is

$$\hat{F}_n(t,\theta) = P_\theta(\hat{T}_n < t),$$

and let

$$\hat{G}_n(t) = \inf\{\hat{F}_n(t,\theta) : \theta \in \Theta_o\},$$

for $-\infty \leq t \leq \infty$. Define

$$H(\hat{T}_n) = 1 - \hat{G}_n(\hat{T}_n(s)).$$

Bahadur(1965,1967,1971) proved the following results in a series of papers. For each θ in $\Theta - \Theta_o$,

$$\lim_{n\to\infty} \frac{1}{n}\log H(\hat{T}_n) = -J(\theta), \tag{3.6}$$

with probability one when θ obtains. Also Berk(1984) showed high-order asymptotic properties of $\hat{T}_n$.

4. Exact Bahadur Slope of QLRT Statistics

Now consider a case where we cannot get the MLE in the whole space Θ but we can have the MLE in the null space Θ_0 when testing the null hypothesis that $\theta \in \Theta_0$. Define

$$T_n^* = \frac{1}{n}\log\frac{L(\theta^*|\underline{x})}{L(\hat{\theta}_o|\underline{x})}, \tag{4.1}$$

and rewrite $\hat{T}_n$ as

$$\hat{T}_n = \frac{1}{n}\log\frac{L(\hat{\theta}|\underline{x})}{L(\hat{\theta}_o|\underline{x})}, \tag{4.2}$$

where $\hat{\theta}$ and $\hat{\theta}_o$ are MLEs of θ under the whole and the null spaces, respectively. And let θ^* be a consistent estimator of the true parameter under the whole parameter space, that is,

$$\theta^* = \theta + o_p(1).$$

Hence T_n^* is a version of the transformed QLRT statistic.

Assumption 4.1. *The density function* $f(x|\cdot)$ *of* X *satisfies*

$$E_\theta \sup_{d \in N(z)} |\log f(X|d)| < \infty \tag{4.3}$$

for any θ, $z \in \Theta$, *where* $N(z)$ *is some neighborhood of* z.

Define

$$F_n^*(t, \theta) = P_\theta(T_n^* < t)$$

and

$$G_n^*(t) = \inf\{F_n^*(t, \theta) : \theta \in \Theta_0\},$$

for $-\infty \leq t \leq \infty$. Define the level attained by T_n^* as

$$H(T_n^*) = 1 - G_n^*(T_n^*(\underline{x})).$$

Then we have the following theorem.

Theorem 4.1. *Under the Assumption* 4.1 *and the Bahadur conditions for the Bahadur optimality of the LRT statistic, the level attained by the QLRT statistic* T_n^* *satisfies*

$$\frac{1}{n} \log H(T_n^*) \longrightarrow -J(\theta), \tag{4.4}$$

in probability for $\theta \in \Theta - \Theta_o$. *Thus tests based on* T_n^* *defined in* (4.1) *are Bahadur optimal as well as is the LRT.*

Proof: By (3.5)

$$\liminf_{n \to \infty} \frac{1}{n} \log H(T_n^*) \geq -J(\theta), \text{ a.s.}$$

Therefore we need to establish the reverse direction of the inequality. Since $\hat{\theta}$ is the MLE and θ^* is not,

$$L(\theta^*|\underline{x}) \leq L(\hat{\theta}|\underline{x}),$$

which implies

$$T_n^* \leq \hat{T}_n.$$

Thus for any t, $-\infty < t < \infty$,

$$P_\theta(T_n^* \geq t) \leq P_\theta(\hat{T}_n \geq t)$$

for $\theta \in \Theta$. It was shown in Bahadur(1971) and Bahadur and Raghavachari (1970) that for any t in the interior of $\{J(\theta) : \theta \in \Theta - \Theta_0\}$

$$\begin{aligned} \limsup_{n \longrightarrow \infty} \frac{1}{n} \log P_\theta(T_n^* \geq t) &\leq \limsup_{n \longrightarrow \infty} \frac{1}{n} \log P_\theta(\hat{T}_n \geq t) \\ &\leq -t. \end{aligned} \tag{4.5}$$

We rewrite the likelihood function as

$$\frac{1}{n} \log L(z) = \frac{1}{n} \sum_{i=1}^{n} \log f(x_i|z).$$

By the SLLN

$$\begin{aligned} \frac{1}{n} \log L(z) &= \frac{1}{n} \sum_{i=1}^{n} \log f(x_i|z) \\ &\xrightarrow{a.s[\theta]} \gamma_\theta(z), \end{aligned}$$

where

$$\gamma_\theta(z) = E_\theta \log f(X|z).$$

Consider a sequence $\{z_n\}$ in some neighborhood $N(z)$ of z for $z \in \Theta$, such that $z_n \longrightarrow z$ as $n \longrightarrow \infty$. Let

$$\gamma_n(z_n) = \frac{1}{n} \sum_{i=1}^{n} \log f(x_i|z_n),$$

$$\bar{\gamma}_n(z) = \frac{1}{n}\sum_{i=1}^{n} \sup_{d \in N(z)} \log f(x_i|d),$$

and

$$\underline{\gamma}_n(z) = \frac{1}{n}\sum_{i=1}^{n} \inf_{d \in N(z)} \log f(x_i|d).$$

Then

$$\underline{\gamma}_n(z) \leq \gamma_n(z_n) \leq \bar{\gamma}_n(z). \tag{4.6}$$

By the SLLN

$$\bar{\gamma}_n(z) \stackrel{a.s[\theta]}{\longrightarrow} E_\theta \sup_{d \in N(z)} \log f(X|d),$$

and

$$\underline{\gamma}_n(z) \stackrel{a.s[\theta]}{\longrightarrow} E_\theta \inf_{d \in N(z)} \log f(X|d).$$

By the Assumption (4.1), and the Dominated Convergence Theorem

$$\begin{aligned} E_\theta \sup_{d \in N(z)} \log f(X|d) \quad &\downarrow \quad E_\theta \log f(X|z) \\ &= \quad \gamma_\theta(z) \end{aligned}$$

and

$$\begin{aligned} E_\theta \inf_{d \in N(z)} \log f(X|d) \quad &\uparrow \quad E_\theta \log f(X|z) \\ &= \quad \gamma_\theta(z), \end{aligned}$$

as $N(z) \downarrow \{z\}$, where it is assumed by Bahadur conditions that

$$\sup_{d \in N(z)} \log f(X|d) \longrightarrow \log f(X|z),$$

and

$$\inf_{d \in N(z)} \log f(X|d) \longrightarrow \log f(X|z),$$

as $N(z) \downarrow \{z\}$. Therefore by the inequality (4.6)

$$\gamma_n(z_n) \xrightarrow{a.s[\theta]} \gamma_\theta(z), \tag{4.7}$$

as $n \longrightarrow \infty$.

Now consider

$$\begin{aligned} \hat{T}_n - T_n^* &= \frac{1}{n}\log\frac{L(\hat{\theta}|\underline{x})}{L(\hat{\theta}_o|\underline{x})} - \frac{1}{n}\log\frac{L(\theta^*|\underline{x})}{L(\hat{\theta}_o|\underline{x})} \\ &= \frac{1}{n}l_n(\hat{\theta}) - \frac{1}{n}l_n(\theta^*) \\ &= \gamma_n(\hat{\theta}) - \gamma_n(\theta^*). \end{aligned}$$

If θ_o is a true parameter, then by (4.7)

$$\begin{aligned} \hat{T}_n - T_n^* &\xrightarrow{p} \gamma_\theta(\theta_o) - \gamma_\theta(\theta_o) \\ &= 0, \end{aligned} \tag{4.8}$$

since both $\hat{\theta}$ and θ^* are consistent estimators of θ_o.

When θ in $\Theta - \Theta_o$ obtains, we have by (4.8)

$$T_n^* \xrightarrow{p} J(\theta), \tag{4.9}$$

since

$$\hat{T}_n \xrightarrow{a.s[\theta]} J(\theta).$$

It follows from (4.5) and (4.9) that

$$\begin{aligned} \limsup_{n\longrightarrow\infty} \frac{1}{n}\log H(T_n^*) &= \limsup_{n\longrightarrow\infty} \frac{1}{n}\log(1 - G_n^*(T_n^*(\underline{x}))) \\ &\leq -J(\theta), \end{aligned}$$

in probability for $\theta \in \Theta - \Theta_o$.

This completes the proof. □

The QLRT based on one-step or two-step approximation of the Newton-Raphson method or the Method of Scoring is Bahadur optimal under the Bahadur conditions and the regularity conditions for the asymptotic distribution of the LRT statistic as long as we have the MLE under the null hypothesis or we have a simple null hypothesis.

5. Bartlett-Adjustable QLRT Statistics

Suppose $\hat{\lambda}$ is the transformed LRT statistic (defined later in this section), whose limiting distribution is a χ^2 distribution. Since Wilks(1938) found the limiting distribution, many statisticians have been worked on its asymptotic distribution in order to improve it; Lawley(1956), Hayakawa(1975,1977), Barndorff-Nielsen and Cox(1984), and Barndorff-Nielsen and Hall(1988).

Consider the case where we cannot get MLE under the whole parameter space but we have MLE under the null parameter space considered in the previous section. The following results can readily apply to cases where we cannot get MLE under both parameter spaces. Let

$$\hat{\lambda} = -2 \log \hat{\Lambda}_n,$$

which can be rewritten as

$$\hat{\lambda} = 2\{l_n(\hat{\theta}) - l_n(\theta)\},$$

where $l_n(\theta)$ is a log-likelihood function.

Define a d-step QLRT statistic $\tilde{\lambda}_d$ as follow:

$$\tilde{\lambda}_d = 2\{l_n(\delta_d) - l_n(\theta)\},$$

where δ_d is a d-step approximation of Newton-Raphson method. Let β be the rate at which the d-step QLRT statistic converges to the LRT statistic, i.e.

$$\tilde{\lambda}_d = \hat{\lambda} + O_p(n^{-\beta}).$$

And we define any consistent estimator T_n of a parameter θ as α-consistent if

$$T_n = \theta + O_p(n^{-\alpha}).$$

Then we have the following result.

Theorem 5.1. *When we iterate the Newton-Raphson approximation with an α-consistent estimator as a starting variable, the rate at which the d-step QLRT statistic converges to the LRT statistic is*

$$\beta = d,$$

that is,

$$\tilde{\lambda}_d = \hat{\lambda} + O_p(n^{-d}),$$

where α is greater than, or equal to $1/2$, and $d = 1, 2, 3, \cdots$.

Proof: Expand $l(\delta_d)$ in Taylor series in $\hat{\theta}$

$$l(\delta_d) = l(\hat{\theta}) + (\delta_d - \hat{\theta})l'(\hat{\theta}) + \frac{1}{2!}(\delta_d - \hat{\theta})^2 l''(\theta_1),$$

where θ_1 lies between δ_d and $\hat{\theta}$ and $l''(\theta_1) = O_p(n)$ by the regularity conditions for asymptotic distribution of the MLE. Since $l'(\hat{\theta}) = 0$ and

$$\delta_d = \hat{\theta} + O_p(n^{-\frac{1}{2}(d+1)})$$

we write

$$l(\delta_d) = l(\hat{\theta}) + O_p(n^{-d}).$$

Therefore we have

$$\begin{aligned}
\tilde{\lambda}_d &= 2\{l(\delta_d) - l(\theta)\} \\
&= 2\{l(\hat{\theta}) + O_p(n^{-d}) - l(\theta)\} \\
&= \hat{\lambda} + O_p(n^{-d}).
\end{aligned}$$

This completes the proof. □

It is of special interest to find out the number of steps required to make the QLRT Bartlett adjustable.

Theorem 5.2. *When we iterate the Newton-Raphson approximation with an α-consistent estimator as a starting variable, the d-step QLRT statistic is Bartlett adjustable, where α is greater than, or equal to $1/2$, and $d = 2, 3, \cdots$.*

Proof: Let w be a statistic whose asymptotic distribution is chi-square with k degrees of freedom and whose expectation can be expanded as

$$\begin{aligned}
Ew &= kb \\
&= k\{1 + \frac{b_1}{n} + O(n^{-\frac{3}{2}})\}.
\end{aligned}$$

Define w' to be a Bartlett adjusted statistic, namely, $w' = (1 + b_1/n)^{-1}w$. Then it is seen that

$$f_{w'}(x) = q_k(x) + O(n^{-\frac{3}{2}}) \tag{5.1}$$

if and only if

$$f_w(x) = (1 - \frac{kb_1}{2n})q_k(x) + \frac{kb_1}{2n}q_{k+2}(x) + O(n^{-\frac{3}{2}}),$$

where $q_k(\cdot)$ is a chi-square density with k degrees of freedom. See Barndorff-Nielsen and Cox(1984) for details.

Because of the Bartlett adjustability of the LRT statistic, its density can be written as (5.1). It is enough to show that the d-step QLRT is Bartlett adjustable for $d = 2$ and $\alpha = 1/2$. When $d = 2$ and $\alpha = 1/2$, the rate of the d-step QLRT statistics relative to the corresponding LRT statistic is $\beta = 2$, that is,

$$\tilde{\lambda}_2 = \hat{\lambda} + O_p(n^{-2}).$$

Therefore up to $O(n^{-\frac{3}{2}})$, $\tilde{\lambda}_2$ and $\hat{\lambda}$ have the same density (5.1). This completes the proof. □

Barndorff-Nielsen and Hall (1988) showed that the Bartlett adjusted LRTs converge to a χ^2 variate up to order $O(n^{-2})$. In other words, there is no $O_p(n^{-\frac{3}{2}})$ term in Edgeworth expansions of the densities of the adjusted LRT.

6. Concluding Remarks

In estimation problems where the MLE is not feasible, we can find asymptotically efficient estimators using some iteration methods. In testing problems where the LRT cannot be applicable, we have used some alternatives such as Wald test and score test, which are neither Bahadur optimal nor Bartlett adjustable. When we can find the MLE under the null parameter space, the suggested QLRT is Bahadur optimal and Bartlett adjustable. We need some more works for the case where the MLEs are not obtainable under both the null and the whole parameter spaces.

REFERENCES

Bahadur, R.R.(1960a). Stochastic comparison of tests. *Annals of Mathematical Statistics*, Vol 31, 276-295.

Bahadur, R.R.(1960b). Asymptotic efficiency of tests and estimates. *Sankya* Vol 22, 229-252.

Bahadur, R.R.(1965). An optimal property of the likelihood ratio statistic, Proceeding of Fifth Berkeley Symposium in Mathematical Statistics and Probability, 13-26.

Bahadur, R.R.(1967). Rate of convergence of estimates and test statistics, *Annals of Mathematical Statistics*, Vol 38, 303-324.

Bahadur, R.R.(1971). Some limits theorems in Statistics, Society for Industrial and Applied Mathematics.

Bahadur, R.R. and Raghavachari, M.(1970). Some asymptotic properties of likelihood ratios and general sample space, Proceeding of Sixth Berkeley Symposium in Mathematical Statistics and Probability, 129-152.

Barndorff-Nielsen, O.E. and Cox, D.R.(1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator, *Journal of Royal Statistical Society*, Series B, 41, 483-495.

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

Berk, R.H.(1984). Stochastic bounds for attained levels, *Journal of Multivariate Analysis*, Vol. 14, 376-389.

Hayakawa, T.(1975). The likelihood ratio criterion for a composite hypothesis under a local alternative, *Biometrika*, 62, 451-460.

Hayakawa, T.(1977). The likelihood ratio criterion and the asymptotic expansion of its distribution, *Annals of Institute of Mathematical Statistics*, 29, 359-378.

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

Lee, K.J. (1992). Unpublished Ph.D dissertation at Rutgers University.

Raghavachari M.(1970). On a theorem of Bahadur on the rate of convergence of test statistics, *Annals of Mathematical Statistics* , 41, 1695-1699.

Wilks, S.S.(1938). The large-sample distribution of the likelihood ratio for testing composite hypothesis, *Annals of Mathematical Statistics*, 9, 60-62.