Ann. Inst. Statist. Math.
Vol. 48, No. 2, 295–314 (1996)

BAHADUR EFFICIENCY AND ROBUSTNESS OF STUDENTIZED SCORE TESTS

XUMING HE[1] AND QI-MAN SHAO[2]

[1] *Department of Statistics, University of Illinois, Champaign, IL 61820, U.S.A.*
[2] *Department of Mathematics, National University of Singapore, Singapore 0511*

(Received July 27, 1994; revised July 14, 1995)

Abstract. We derive the exact Bahadur slopes of studentized score tests for a simple null hypothesis in a one-parameter family of distributions. The Student's t-test is included as a special case for which a recent result of Rukhin (1993, *Sankhyā Ser. A*, **55**, 159–163) was improved upon. It is shown that locally optimal Bahadur efficiency for one-sample location models with a known or estimated scale parameter is attained within the class of studentized score tests. The studentized test has an asymptotic null distribution free of the scale parameter, and the optimality of likelihood scores does not depend on the existence of a moment generating function. We also consider the influence function and breakdown point of such tests as part of our robustness investigation. The influence of any studentized score test is bounded from above, indicating certain degree of robustness of validity, but a bounded score function is needed to cap the influence from below and to ensure a high power breakdown point. We find that the standard Huber-type score tests are not only locally minimax in Bahadur efficiency, but also very competitive in global efficiency at a variety of location models.

Key words and phrases: Bahadur slope, efficiency, influence function, score test.

1. Introduction

A statistical test making a binary decision is subject to two types of errors, commonly known as type I and type II errors. A variety of definitions of asymptotic efficiency of tests have been studied in the literature to account for the two-sided risk. Some are based on the power of a test for a fixed level of significance. Others place emphasis on the size of the type I error while keeping the type II error in check. In this paper, we use Bahadur efficiency as a measure of test performance, and consider testing the null hypothesis of $H_0 : \theta = 0$ versus the alternative hypothesis of $H_1 : \theta > 0$ based on a random sample from a one-parameter family of distributions $\{f_\theta, \theta \in R\}$. Special emphasis is given to location models where $f_\theta(x) = s_0^{-1} f((x - \theta)/s_0)$ for some known or unknown but fixed scale parameter

s_0. With some trivial modifications, all results in this paper hold for the two-sided alternative hypothesis.

Bahadur efficiency is closely related to the observed significance level or the so-called p-value of a test. For each $n = 1, 2, \ldots$, let T_n be an extended real-valued function such that T_n is measurable in the σ-field generated by $X_1, X_2, \ldots, X_n$. Assume that large values of T_n are significant, and that T_n has a null distribution F_n, then the level attained by T_n is given by

$$P_n = 1 - F_n(T_n). \tag{1.1}$$

We say that the test based on T_n has Bahadur slope $c(\theta)$ if

$$\lim_{n\to\infty} n^{-1} \log P_n = -c(\theta)/2 \quad \text{a.s.} \tag{1.2}$$

where the sample $X_1, X_2, \ldots, X_n$ is drawn from f_θ. More generally, we define

$$\bar{c}(\theta) = -2 \liminf_{n\to\infty} n^{-1} \log P_n,$$

and

$$\underline{c}(\theta) = -2 \limsup_{n\to\infty} n^{-1} \log P_n$$

as upper and lower Bahadur slopes. They may be used when existence of the limit in (1.2) is not assured. A somewhat artificial example is to take T_n to be the sample mean for some n but the sample median for others.

The null distribution of the p-value P_n was obtained by Bahadur (1960) for some special cases and by Lambert and Hall (1982) for more general settings. In typical cases, the p-value is found to be asymptotically lognormal under any alternative distribution. The Bahadur slope which measures the exponential rate at which the p-value approaches zero is twice the mean of the limiting lognormal distribution.

Bahadur efficiency refers to the size of Bahadur slope. It indicates how fast the type I error can be asymptotically eliminated for a specification of power at each given alternative. Naturally, this depends on the distance between the null distribution and the alternative distribution. The larger the distance, the easier it becomes to distinguish one from the other. Bahadur (1965, 1971) used the Kullback-Leibler information number to establish an upper bound on Bahadur slopes, that is,

$$\underline{c}(\theta) \le \bar{c}(\theta) \le 2E_\theta \log\{f_\theta(X)/f_0(X)\}. \tag{1.3}$$

It was also shown that if $c(\theta)$ exists, the optimal Bahadur slope is attained by the likelihood ratio test under general regularity conditions.

The likelihood ratio test has an asymptotic chi-square null distribution, but an explicit form of the test statistic is available only at some special models. In general, it may be computationally inconvenient and suffers from non-robustness against model misspecification or occurrence of outliers. The Rao-type score tests often become attractive alternatives. Akritas and Kourouklis (1988) showed that

$$T_n = n^{-1/2} \sum_{i=1}^{n} f'(X_i)/f(X_i) \tag{1.4}$$

is locally optimal in Bahadur efficiency. In fact, this score test is asymptotically equivalent to the likelihood ratio test under the null distribution and has optimal local efficiency in almost every definition considered in the literature.

However, the formulation of (1.4) makes an implicit assumption that the variance of $\psi^*(X) = f'(X)/f(X)$ is known and finite. In practice, the variance is typically unknown. Consider the Gaussian model where (1.4) reduces to a test based on the sample mean $\bar{X}_n$, a more realistic test would be based on the studentized mean, or the popular t-test. This seemingly routine standardization usually has no effect on the asymptotic distributions of the test statistic under the null or contiguous alternative distributions, but it poses a challenging problem for the determination of exact Bahadur efficiency in general. In the rest of the paper, we consider the class of studentized score tests in the form of

$$T_n = \frac{\sum_{i=1}^{n} \psi(X_i)}{\sqrt{\sum_{i=1}^{n} \psi^2(X_i)}} \tag{1.5}$$

where ψ is a score function such that $E_\theta \psi(X) \geq 0$ where the equality holds if and only if $\theta = 0$. This condition is necessary to rule out inconsistent tests for the hypothesis testing problem we consider. We also consider the use of an auxiliary scale estimate to achieve scale invariance of the test in (2.5) below. For a related work using approximate pivots from M-estimators, see Lloyd (1994).

If $\psi(x) = x$, (1.5) is just a monotone transformation of the conventional t-test statistic whose Bahadur efficiency was derived by Rukhin (1993) under certain conditions. As a matter of fact, Efron (1969) showed that this equivalent form has an advantage over the conventional one: the critical values are extremely stable across a wide range of sample sizes n so that the asymptotic approximation works better. In general, the use of studentization in (1.5) or (2.5) avoids the need to estimate the asymptotic variance of T_n as the resulting distribution is asymptotically free of the scale parameter.

Mainly due to the difficulty in evaluating the exact Bahadur slopes, some authors have been resorting to approximate Bahadur slopes by using the limiting null distribution instead of the exact distribution of T_n. In some contexts, the approximate Bahadur slopes are the more relevant quantities, see Lambert (1981). However, the approximate Bahadur efficiency suffers the lack of invariance to monotone transformations of the test statistic and could be sometimes misleading (cf. Groeneboom and Oosterhoff (1977)). Equivalence of the local limit of exact Bahadur efficiency and Pitman efficiency is shown in Kallenberg and Ledwina (1987) for a class of weakly continuous statistical functionals. The functionals corresponding to the studentized score tests of (1.5), however, do not have the continuity required.

A recent work of Shao (1994) on large deviation of self-normalized sums provided us with the necessary technicality to evaluate the exact Bahadur slope of all score tests in (1.5) at any univariate model. In Section 2, we derive the Bahadur slopes and show that under mild conditions the likelihood score $\psi(x) = f'(x)/f(x)$ is indeed locally optimal in Bahadur efficiency. At location models, the same local optimality remains valid if an auxiliary estimate of the scale parameter is needed in (1.5) to achieve scale invariance of the test. In fact, the Bahadur efficiency of any score function is locally equivalent to the asymptotic efficiency of the corresponding M-estimator. By a result of Huber (1964), a robust test of locally minimax Bahadur efficiency can be easily constructed. Robustness of tests in contamination neighborhood is discussed in Section 3, based on the standard notion of influence function and breakdown point of tests. The studentization results in an automatic boundedness of the influence function from above. But a bounded score function is still necessary for robustness of power. For a variety of models, including logistic, normal and t-distributions, the Huber's score function is found to be highly competitive in Bahadur efficiency with the locally optimal test and actually does better for modest to large alternative location parameters. The t-test is less attractive at non-Gaussian models. The technicality needed to incorporate the use of a scale estimate in the derivation of local Bahadur slopes is rather complicated, and therefore deferred to Section 4.

2. Locally optimal score test

In this section, we begin with any parametric model for a random variable or vector X with probability density function $\{f_\theta, \theta \in R\}$. The null hypothesis is the same as in the introduction. The calculation of Bahadur slope applies to the studentized score test in general one-parameter models including the location model. Whenever we write f or E without subscript, we mean the density function or expectation under the null hypothesis. We impose minimal regularity conditions on the model distributions.

(R1) The Fisher information $I_0 = E(f'(X)/f(X))^2$ is finite.

(R2) The Kullback-Leibler information number $K(\theta) = E_\theta \log\{f_\theta(X)/f(X)\}$ is finite and locally quadratic $K(\theta) = \frac{1}{2} I_0 \theta^2 + o(\theta^2)$ as $\theta \to 0$.

(R3) $P\{\psi(X) = 0\} = 0$, $E\psi(X) = 0$ and $E_\theta \psi(X) > 0$ for any $\theta > 0$.

(R4) $\gamma = \lim_{\theta \downarrow 0} E_\theta \psi(X)/\theta$ exists.

It can be seen from the following derivations that if $E_\theta \psi^2(X)$ is infinite, the Bahadur slope at θ is zero. Therefore, we may restrict ourselves to score functions with finite second moments.

THEOREM 2.1. *Under the condition* (R3), *the Bahadur slope of the studentized score test* T_n *of* (1.5) *is given by*

$$c_\psi(\theta) = -2\log\{\sup_{c \geq 0} \inf_{t \geq 0} E e^{t(2c\psi(X) - a(\theta)(c^2 + \psi^2(X)))}\} \tag{2.1}$$

where $a(\theta) = E_\theta \psi(X)/\sqrt{E_\theta \psi^2(X)}$. *Under the additional condition* (R4), *the local Bahadur slope is determined by*

$$\lim_{\theta \downarrow 0} c_\psi(\theta)/\theta^2 = \gamma^2 / E\psi^2(X). \tag{2.2}$$

The statement (2.1) follows from Theorem 1.1 of Shao (1994) and Theorem 7.2 of Bahadur ((1971), p. 27). The continuity of $w(a) = \sup_{c\geq 0}\inf_{t\geq 0} Ee^{t(2c\psi(X)-a(c^2+\psi^2(X)))}$ in $a > 0$ was established in Shao (1994). The local result (2.2) can be derived from Theorem 3.1 of Shao (1994) using similar arguments to those in our proof of Theorem 2.3 below. We omit the details.

Under a stronger condition that the moment generating function of (X, X^2) exists in a neighborhood of the origin, Rukhin (1993) obtained the Bahadur slope of the t-test as

$$c_T(\theta) = -2\sup_{c\geq 0}\inf_{t_1 t_2}\log Ee^{a(\theta)ct_1+c^2t_2-t_1X-t_2X^2}.$$

This can be shown to be equivalent to (2.1) for $\psi(x) = x$. Rukhin (1993) argued further that the t-test is never globally optimal in Bahadur slopes. Theorem 2.1 in the present paper is valid without the existence of a moment generating function, thus generalizing the results of Rukhin (1993).

The exact Bahadur slope of (2.1) is involved with a double optimization process whose analytic solutions are possible only with some special choices of f and ψ. However, the bivariate function in c and t is usually well behaved and a numerical solution is not difficult to obtain, see the end of Section 3 for more details.

The local Bahadur efficiency, however, is quite simple. It is equivalent to the asymptotic variance of the M-estimator of θ determined by $\sum_i \psi_\theta(X_i) = 0$. If $\{f_\theta\}$ is a location family of distributions such that $f_\theta(x) = f(x-\theta)$, then the score function $\psi^*(x) = f'(x)/f(x)$ attains its highest Bahadur slope locally.

THEOREM 2.2. *Consider a location family of distributions* $\{f(x-\theta), \theta \in R\}$. *If* $E_\theta\psi^*(X)$ *is differentiable under the integral sign at* $\theta = 0$ *where* $\psi^*(x) = f'(x)/f(x)$, *then under the conditions* (R1)–(R4),

$$\lim_{\theta\downarrow 0} c_{\psi^*}(\theta)/K(\theta) = 2, \tag{2.3}$$

and the likelihood score test is locally optimal in Bahadur efficiency.

Since the limit in (2.2) for ψ^* becomes the Fisher information I_0, the result follows immediately from (R2) and the upper bound (1.3).

The local optimality in Theorem 2.2 does not require the existence of a moment generating function of $\psi^*(X)$. Akritas and Kourouklis ((1988), p. 191) provided an interesting counterexample for score tests without studentization.

As with other criteria of global test efficiency, this locally optimal score test does not ensure good Bahadur efficiency at all alternatives. Consider the double exponential distribution with mean θ, that is, $f_\theta(x) = \frac{1}{2}e^{-|x-\theta|}$. The locally optimal score test of Theorem 2.2 is the equivalence of the well-known Sign-test whose Bahadur slope is given by

$$c(\theta) = 2\{\log 2 + F(\theta)\log F(\theta) + (1-F(\theta))\log(1-F(\theta))\} \tag{2.4}$$

which has a maximum value of $2\log 2$, see Bahadur ((1971), p. 25). A straightforward calculation shows that $K(\theta) = \theta + e^{-\theta} - 1$. The upper bound of (1.3) increases linearly (to infinity) in θ.

A pure location model is rare in practice. In the presence of an unknown scale parameter, the test of (1.5) is not scale invariant unless ψ is either invariant or equivariant under scale multiplication. One common practice is to standardize the observations and use

$$T_n^s = \frac{\sum_{i=1}^n \psi(X_i/s_n)}{\sqrt{\sum_{i=1}^n \psi^2(X_i/s_n)}} \tag{2.5}$$

where s_n is a preliminary scale estimate (which is required to be scale equivariant). Usually, the asymptotic null distribution of the test statistic does not change so long as s_n is consistent.

When the scale parameter is unknown, there are two forms of hypotheses that we may consider. One is to allow arbitrary scale parameters in both the null and alternative hypotheses. The other is to keep the scale parameter fixed (but unknown) in the parameter space. The latter implies that the location shift is the only difference between the null and alternative distributions. We take on the latter for relative simplicity in the derivation of the exact local Bahadur slope for T_n^s. To this end, our model distribution can be written as $\frac{1}{s_0}f(\frac{x-\theta}{s_0})$ for some $s_0 > 0$, where f is a known density symmetric about zero. We assume that for any $\delta > 0$, there exists $\eta \in (0,1)$ such that

$$P_\theta(|s_n - s_0| > \delta) \le \eta^n \tag{2.6}$$

for all n. Furthermore, we assume the following conditions for the score function. Note that E without subscript denotes the expectation at $\theta = 0$.

(R5) For any $\frac{1}{2} < a < 2$, $|\psi(xa)| \le K(1 + |\psi(x)|)$ for some $K < \infty$.

(R6) There exists $\eta_0 > 0$ such that for any $0 < \delta < \eta_0$ and $1-\eta_0/2 \le s \le 1+2\eta_0$ and $0 \le \theta < \eta_0$, $E_\theta \sup_{1\le t\le 1+\delta}(\psi(sX) - \psi(stX))^2 \le K\delta^\alpha$ for some $\alpha > 0$ and $K > 0$.

Throughout the article, we assume without loss of generality that $s_0 = 1$ and $\eta_0 = 1$.

LEMMA 2.1. *If s_n satisfies* (2.6) *and ψ satisfies $E\psi^2(X) < \infty$ in addition to* (R3), (R5) *and* (R6), *then*

$$\lim_{n\to\infty} x_n^{-2} \log P(T_n^s \ge x_n) = -1/2 \tag{2.7}$$

for any sequence of positive numbers $\{x_n\}$ with $\sqrt{\log n} \le x_n = o(\sqrt{n})$. Furthermore, for any $0 < \epsilon < 1$, there exist δ and N such that

$$e^{-(1+\varepsilon)x^2/2} \le P(T_n^s \ge x) \le e^{-(1-\varepsilon)x^2/2} \tag{2.8}$$

for every $n > N$ and $\sqrt{\log n} \le x \le \delta\sqrt{n}$.

The proof of Lemma 2.1 is provided in Section 4. With the help of (2.8), we obtain

THEOREM 2.3. *Suppose that $E_\theta\psi^2(X) < \infty$ for θ in a right neighborhood of zero. Under the conditions* (R3), (R4), (R5) *and* (R6), *the lower and upper Bahadur slopes of T_n^s satisfies*

$$\lim_{\theta\downarrow 0}\frac{\bar{c}(\theta)}{\theta^2} = \lim_{\theta\downarrow 0}\frac{\underline{c}(\theta)}{\theta^2} = \gamma^2\{E\psi^2(X)\}^{-1}. \tag{2.9}$$

If $\psi(X)$ does not have a finite second moment, (2.9) becomes a trivial identity. Theorem 2.3 implies that locally the score test T_n^s is as good as T_n (with s_0 known) in terms of Bahadur efficiency. Note that Lemma 2.1 does not necessarily guarantee the existence of the Bahadur slope (1.2). If it does, it has the same limit as in (2.2).

The conditions (R5) and (R6) can be easily verified for a wide variety of commonly used model distributions and score functions ψ. Jump discontinuities in ψ are also allowed. A robust scale estimate such as $s_n = median\{|X_i - median(X_j)|\}$ times a normalizing constant may be used for any location-scale model. It can be verified that this choice of s_n satisfies the large deviation requirement of (2.6).

PROOF OF THEOREM 2.3. We first show that

$$\frac{\sum_{i=1}^n \psi^2(X_i/s_n)}{n} \to E_\theta\psi^2(X) \quad \text{a.s. as} \quad n\to\infty. \tag{2.10}$$

For $1/2 < s < 2$ and any $A > 1$, we have, by (R5),

$$\begin{aligned}
|\psi^2(sX) - \psi^2(X)| &\le 2K|\psi(sX) - \psi(X)|(1 + |\psi(X)|) \\
&\le 2K(1+A)|\psi(sX) - \psi(X)| \\
&\quad + 2K|\psi(sX) - \psi(X)|(1 + |\psi(X)|)I_{\{|\psi(X)|>A\}} \\
&\le 2K(1+A)|\psi(sX) - \psi(X)| \\
&\quad + 4K^2(1 + |\psi(X)|)^2 I_{\{|\psi(X)|>A\}},
\end{aligned}$$

and then, by (R6),

$$\begin{aligned}
&\limsup_{\delta\downarrow 0} E_\theta \sup_{(1-\delta)\le s\le(1+\delta)} |\psi^2(sX) - \psi^2(X)| \\
&\quad\le 2K(1+A)\limsup_{\delta\downarrow 0} E_\theta \sup_{(1-\delta)\le s\le(1+\delta)} |\psi(sX) - \psi(X)| \\
&\qquad + 4K^2 E_\theta(1 + |\psi(X)|)^2 I_{\{|\psi(X)|>A\}} \\
&\quad= 4K^2 E_\theta(1 + |\psi(X)|)^2 I_{\{|\psi(X)|>A\}}.
\end{aligned}$$

Letting $A \to \infty$ leads to

$$\lim_{\delta\downarrow 0} E_\theta \sup_{(1-\delta)\le s\le(1+\delta)} |\psi^2(sX) - \psi^2(X)| = 0. \tag{2.11}$$

On the other hand, (2.6) implies $s_n \to 1$ almost surely. Therefore for any $\delta > 0$, it holds for sufficiently large n that

$$\frac{1}{n}\sum_{i=1}^{n}|\psi^2(X_i/s_n)-\psi^2(X_i)| \le \frac{1}{n}\sum_{i=1}^{n}\sup_{(1-\delta)\le s\le(1+\delta)}|\psi^2(sX_i)-\psi^2(X_i)|$$

which converges to $E_\theta \sup_{(1-\delta)\le s\le(1+\delta)}|\psi^2(sX)-\psi^2(X)|$ as $n\to\infty$. Letting $\delta\to 0$ and by (2.11), we have

$$\frac{1}{n}\sum_{i=1}^{n}|\psi^2(X_i/s_n)-\psi^2(X_i)| \to 0 \quad \text{a.s.}$$

By the law of large numbers applied to $\frac{1}{n}\sum_{i=1}^{n}\psi^2(X_i)$, we obtain (2.10).

Similarly, we have

$$\frac{\sum_{i=1}^{n}\psi(X_i/s_n)}{n} \to E_\theta\psi(X) \quad \text{a.s. as} \quad n\to\infty,$$

and therefore,

$$\frac{\sum_{i=1}^{n}\psi(X_i/s_n)}{\sqrt{n\sum_{i=1}^{n}\psi^2(X_i/s_n)}} \to a(\theta) \quad \text{a.s. as} \quad n\to\infty, \tag{2.12}$$

where $a(\theta)=E_\theta\psi(X)/\sqrt{E_\theta\psi^2(X)}$.

Similar to Theorem 7.2 of Bahadur (1971), we have

$$\underline{c}(\theta) \ge -2\lim_{\epsilon\downarrow 0}\limsup_{n\to\infty} n^{-1}\log P(T_n^s \ge (1-\epsilon)a(\theta)\sqrt{n}), \tag{2.13}$$

and

$$\bar{c}(\theta) \le -2\lim_{\epsilon\downarrow 0}\liminf_{n\to\infty} n^{-1}\log P(T_n^s \ge (1+\epsilon)a(\theta)\sqrt{n}). \tag{2.14}$$

By (2.8) and the fact that $\lim_{\theta\downarrow 0}a(\theta)=0$, we have, for any $-1/4<\eta<1/4$

$$\begin{aligned}\lim_{\theta\downarrow 0}\theta^{-2}\Big(-2\log\lim_{n\to\infty}P(T_n^s \ge a(\theta)(1+\eta)\sqrt{n})^{1/n}\Big) \\ = \lim_{\theta\downarrow 0}(1+\eta)^2a^2(\theta)/\theta^2 = (1+\eta)^2\gamma^2/E\psi^2(X).\end{aligned} \tag{2.15}$$

Taken together, (2.12)–(2.15) imply (2.9). The proof is then complete.

3. Robustness of score tests

An optimal test at a particular model does not imply any goodness of the test at a different, even though neighboring, model, unless there is a built-in robustness property in the test. Parametric models are rarely meant to be exact in practice, so an optimal test without being robust isn't worth as much as it sounds. For instance, the t-test would have Bahadur slope equal to zero if the random variable does not have a finite second moment, even though it remains a consistent test so long as the mean exists.

Robustness of a test can be studied from several different perspectives. In this section, we consider the influence function and breakdown points for the score test.

Hampel (1974) defined the influence function of a statistical functional $T(\cdot)$ at F as

$$IF(x; T, F) = \frac{d}{d\epsilon} T((1-\epsilon)F + \epsilon\delta_x) \mid_{\epsilon=0} \tag{3.1}$$

where δ_x denotes the point mass distribution at the point x. This has been used as a standard measure of local robustness for parameter estimators. It provides a heuristic measure as to how much the functional can be influenced by infinitesimal contamination at a single point x. Several variants of the influence function have been proposed to study the robustness of a test. For example, the influence can be measured on the level, power or p-value of the test. Lambert (1981) applied Hampel's influence function to the transformed p-values and showed that the influence of the t-test is bounded from above, but not from below. It indicates that an outlier may inflate the t-statistic, but cannot pull the p-value arbitrarily close to zero under the normal alternative. Lambert's definition of influence function, based on the Bahadur slope at contaminated alternative distributions, makes a direct computation highly nontrivial. Hampel *et al.* (1986) proposed to use the influence function for a transformed test statistic $U^{-1}(T)$ where U^{-1} is the inverse function of $U(\theta) = E_\theta T$. The transformation from T to $U^{-1}(T)$ is to make the resulting functional on the scale of the original parameter θ so that the influence functions of different test statistics are directly comparable. They showed that if the p-value depends on data through the test statistic T_n alone, the influence function of Lambert is proportional to the influence function of $U^{-1}(T)$. If the Bahadur slope is non-constant in θ, the qualitative robustness (ie. the boundedness of influence function) of one implies the other. Therefore, we follow Hampel *et al.* (1986) to derive the influence function of the score test as follow.

THEOREM 3.1. *The influence function of the score test* (1.5) *at the distribution* F_θ *is given by*

$$IF(x; T, F_\theta) = \frac{E_\theta\psi(X)}{h'(\theta)(E_\theta\psi^2(X))^2}(2\psi(x)E_\theta\psi^2(X) - \psi^2(x)) - \frac{h(\theta)}{h'(\theta)} \tag{3.2}$$

where $h(\theta) = (E_\theta\psi(X))^2/E_\theta\psi^2(X)$.

PROOF. Since the influence function of the test is invariant under one-to-one transformations, we work with $h^-(T^2)$, where h^- is the inverse function of h and $T(F) = (E_F\psi(X))^2/E_F\psi^2(X)$ for any distribution F.

Let $H = (1-\epsilon)F_\theta + \epsilon\delta_x$. Direct calculations show

$$(E_H\psi(X))^2 = (E_\theta\psi)^2 + 2\epsilon E_\theta\psi(\psi(x) - E_\theta\psi) + O(\epsilon^2),$$

and

$$\{E_H\psi^2(X)\}^{-1} = (E_\theta\psi^2)^{-1} - \epsilon(\psi^2(x) - E_\theta\psi^2) + O(\epsilon^2).$$

Taking derivative of $h^-(T^2(H))$ with respect to ϵ at $\epsilon = 0$, we obtain (3.2) after suitable rearrangements of terms.

If the scale estimate s_n is representable by a function $S(F)$ which is Fisher consistent and has an influence function, then (3.2) also holds for the test statistic T_n^s.

The score test has a bounded influence function for each bounded ψ. More interesting is the fact that if ψ is unbounded, the influence function is bounded from above or below at F_θ depending on the sign of $E_\theta\psi(X)/h'(\theta)$. In typical applications with $\frac{d}{d\theta}E_\theta\psi(X) > 0$ at $\theta = 0$, the influence is always bounded above at the null distribution. In this sense, every studentized score test has some robustness of validity: the effect of outliers is limited under the null distribution. This is another advantage from studentization. The non-standardized score test has unbounded influence in either direction if ψ is unbounded.

The preceding discussions also enable us to arrive at a score test of locally minimax Bahadur efficiency over contamination neighborhood in the simple location problems. For simplicity, we consider the Gaussian location model and assume ψ to be continuously differentiable. Since $\lim_{\theta\downarrow 0} c(\theta)/\theta^2 = (\int \psi'(x)dF)^2/\int \psi^2(x)dF$, the locally minimax Bahadur efficient score test maximizes

$$\inf\left\{\left(\int \psi'(x)dF\right)^2 \Big/ \int \psi^2(x)dF, F = (1-\epsilon)F_0 + \epsilon G, \text{ any } G\right\}.$$

This is exactly the same problem worked out by Huber (1964) for minimax variance M-estimators. The solution is the well known Huber's score function

$$\psi_c(x) = median\{x, c, -c\} \tag{3.3}$$

where c depends on the contamination size ϵ. With no contamination, the minimax test is the t-test. If the contamination is allowed to be as large as 50%, the Sign-test is called for. Huber (1965) constructed the same minimax test from the likelihood ratio approach, using local power as optimality criterion.

Huber's score function also guarantees a good global stability as evidenced by its high breakdown point. The power and level breakdown of a test statistic was studied in He *et al.* (1990). The level breakdown point gives the smallest amount of contamination of the null distribution under which the test statistic will favor any alternative over the null. The power breakdown point is the least fraction of contamination needed to render the test inconsistent at every alternative. High level and power breakdown points are desirable for a test to be robust against multiple outliers. The breakdown property of the non-standardized score tests is given in Theorem 3.1 of He *et al.* (1990). The following proposition stated for T_n can be proved in the same way. If s_n is a scale estimate with breakdown point 1/2, the same is true for T_n^s.

PROPOSITION 3.1. *The level breakdown point of any score test is* 1. *The power breakdown point is* 1/2 *if* ψ *is bounded, but zero if* ψ *is unbounded.*

Table 1. Bahadur slopes of selected score tests.

θ	0.2	0.4	0.6	1.0	2.0	3.0
Huber	0.03	0.14	0.31	0.70	1.77	3.04
t-test	0.04	0.15	0.31	0.69	1.61	2.30
bound	0.04	0.16	0.36	1.00	4.00	9.00

(a) Normal $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

θ	0.2	0.4	0.6	1.0	2.0	3.0
Huber	0.01	0.05	0.11	0.29	0.92	1.74
t-test	0.00	0.04	0.10	0.28	0.85	1.41
$\psi*$	0.01	0.05	0.11	0.29	0.86	1.46
bound	0.01	0.05	0.12	0.33	1.25	2.63

(b) Logistic $f(x) = \frac{e^{-x}}{(1+e^{-x})^2}$.

θ	0.2	0.4	0.6	1.0	2.0	3.0
Huber	0.03	0.11	0.25	0.63	1.72	2.74
t-test	0.02	0.09	0.19	0.48	1.29	1.97
$\psi*$	0.04	0.15	0.30	0.63	1.28	1.56
bound	0.04	0.16	0.35	0.92	2.95	5.13

(c) t-distribution with 3 degrees of freedom.

θ	0.2	0.4	0.6	1.0	2.0	3.0
Huber	0.01	0.03	0.07	0.19	0.60	1.04
$\psi*$	0.02	0.07	0.14	0.31	0.58	0.71
bound	0.02	0.08	0.17	0.45	1.39	2.36

(d) Cauchy $f(x) = \frac{1}{\pi(1+x^2)}$.

Under general conditions, 1/2 is the highest possible power breakdown point achievable for the hypothesis testing problem we consider. As observed in He *et al.* (1990), the score test has a level breakdown point of one versus 1/2 for tests based on parameter estimates.

To examine their global Bahadur efficiencies, we use (2.1) to compute the Bahadur slopes at several common location models for the locally optimal score ψ^*, the t-test, and the Huber-score (3.3) with constant $c = 3$. They are all compared with the optimal slopes at each alternative. Table 1 gives some of the Bahadur slopes at normal, logistic, Cauchy and Student distribution with 3 degrees of freedom. The t-test is not included at the Cauchy model, since it has zero slope at all alternatives.

Numerical calculations for the double optimization of (2.1) are rather straightforward for the range of θ we consider. We found that the inner *inf* function increases from 0 to a maximum and then drops to 0 again as c moves away from zero. For each value of c, the function of t can be minimized at its critical value.

However, for very small or very large values of θ (which correspond to $h(\theta)$ close to 0 or 1), high precision is needed in the calculation to ensure a reasonable relative error.

Table 1 indicates that the locally optimal score is usually close to optimal for alternatives between 0 and 0.4. When θ gets larger, the Bahadur efficiency of the Huber-score increases faster. In fact, the Huber-score is highly competitive at all alternatives and all models considered, in consistency with its robustness property. The t-test is less attractive at non-Gaussian models. It is also worth noting that even at the normal model, the Huber-score outperforms the t-test for $\theta \geq 1$.

4. Proof of Lemma 2.1

PROOF OF LEMMA 2.1. To prove (2.7), we need to show

$$\limsup_{n\to\infty} x_n^{-2} \log P(T_n^s \geq x_n) \leq -1/2, \tag{4.1}$$

and

$$\liminf_{n\to\infty} x_n^{-2} \log P(T_n^s \geq x_n) \geq -1/2. \tag{4.2}$$

For any $0 < \varepsilon < \min(1, \sqrt{E\psi^2(X)})/4$, by (R6), there exists $0 < \delta < 1/4$ such that

$$\begin{aligned}(1-\varepsilon/2)E\psi^2(X) &\leq E \inf_{|s-1|\leq\delta} \psi^2(sX) \\ &\leq E \sup_{|s-1|\leq\delta} \psi^2(sX) \leq (1+\varepsilon/2)E\psi^2(X).\end{aligned} \tag{4.3}$$

By (2.6), we have

$$\begin{aligned}P(T_n^s \geq x_n) &\leq P(T_n^s \geq x_n, |s_n - 1| \leq \delta/2) + P(|s_n - 1| \geq \delta/2) \\ &\leq P(T_n^s \geq x_n, |s_n - 1| \leq \delta/2) + \eta^n.\end{aligned}$$

Let $z_n^2 = n/x_n^2$. We then have

$$\begin{aligned}&P(T_n^s \geq x_n, |s_n - 1| \leq \delta/2) \\ &\leq P\left(\sup_{|s-1|\leq\delta/2} \frac{\sum_{i=1}^n \psi(X_i/s)}{\sqrt{\sum_{i=1}^n \psi^2(X_i/s)}} \geq x_n\right) \\ &\leq P\left(\sup_{|s-1|\leq\delta} \frac{\sum_{i=1}^n \psi(sX_i)}{\sqrt{\sum_{i=1}^n \psi^2(sX_i)}} \geq x_n\right) \\ &\leq P\left(\sup_{|s-1|\leq\delta} \frac{\sum_{i=1}^n \psi(sX_i)I\{|\psi(X_i)| \leq z_n\}}{\sqrt{\sum_{i=1}^n \psi^2(sX_i)}} \geq (1-\varepsilon)x_n\right) \\ &\quad + P\left(\sup_{|s-1|\leq\delta} \frac{\sum_{i=1}^n \psi(sX_i)I\{|\psi(X_i)| > z_n\}}{\sqrt{\sum_{i=1}^n \psi^2(sX_i)}} \geq \varepsilon x_n\right) \\ &\leq P\left(\sup_{|s-1|\leq\delta} \frac{\sum_{i=1}^n \psi(sX_i)I\{|\psi(X_i)| \leq z_n\}}{\sqrt{nE\psi^2(X)}} \geq (1-\varepsilon)^2 x_n\right)\end{aligned}$$

$$
\begin{aligned}
&+P\left(\inf_{|s-1|\le\delta}\sum_{i=1}^n \psi^2(sX_i)\le n(1-\varepsilon)^2E\psi^2(X)\right)\\
&+P\left(\sum_{i=1}^n I\{|\psi(X_i)|>z_n\}\ge\varepsilon^2x_n^2\right)\\
\le\ & P\left(\sup_{|s-1|\le\delta}\sum_{i=1}^n \psi(sX_i)I\{|\psi(X_i)|\le z_n\}\ge(1-\varepsilon)^2x_n\sqrt{nE\psi^2(X)}\right)\\
&+P\left(\sum_{i=1}^n\inf_{|s-1|\le\delta}\psi^2(sX_i)\le n(1-\varepsilon)^2E\psi^2(X)\right)\\
&+P\left(\sum_{i=1}^n I\{|\psi(X_i)|>z_n\}\ge\varepsilon^2x_n^2\right)\\
:=\ & I_1+I_2+I_3.
\end{aligned}
$$

Note that $\sum_{i=1}^n I\{|\psi(X_i)|>z_n\}$ has a binomial distribution, and for any binomial random variable $B(n,p)$,

$$
P(B(n,p)\ge x)\le\left(\frac{enp}{x}\right)^x,\qquad x>0,
$$

it then follows that

$$
\begin{aligned}
I_3&\le\left(\frac{enP(|\psi(X)|>z_n)}{\varepsilon^2x_n^2}\right)^{\varepsilon^2x_n^2}\\
&\le\left(\frac{enE\psi^2(X)I\{\psi(X)|>z_n\}}{z_n^2\varepsilon^2x_n^2}\right)^{\varepsilon^2x_n^2}\\
&\le\left(\frac{eE\psi^2(X)I\{\psi(X)|>z_n\}}{\varepsilon^2}\right)^{\varepsilon^2x_n^2}\le(\varepsilon^{-2}o(1))^{\varepsilon^2x_n^2}.
\end{aligned}
$$

As to I_2, let $\xi_i=\inf_{|s-1|\le\delta}\psi^2(sX_i)$, and $\xi=\inf_{|s-1|\le\delta}\psi^2(sX)$. By (4.3),

$$
\begin{aligned}
I_2&\le P\left(\sum_{i=1}^n(E\xi_i-\xi_i)\ge\varepsilon nE\psi^2(X)\right)\\
&\le\left(\inf_{t>0}e^{-t\varepsilon E\psi^2(X)+tE\xi}Ee^{-t\xi}\right)^n.
\end{aligned}
$$

Take $A>1$ such that $E\psi^2(X)I\{|\psi(X)|>A\}\le\frac{\varepsilon}{4}E\psi^2(X)$. From the elementary inequality

$$
e^x\le1+x+2\min(x^2,|x|),
$$

for any $x\le0$, it follows that for every $t>0$,

$$
\begin{aligned}
Ee^{-t\xi}&\le1-tE\xi+2E\min(t^2\xi^2,t\xi)\\
&\le1-tE\xi+2tE\xi I\{|\psi(X)|>A\}+2t^2E\xi^2I\{|\psi(X)|\le A\}\\
&\le1-tE\xi+2tE\psi^2(X)I\{|\psi(X)|>A\}+2t^2A^4\\
&\le\exp(-tE\xi+t\varepsilon E\psi^2(X)/2+2t^2A^4).
\end{aligned}
$$

Let $t = \varepsilon E\psi^2(X)/(8A^4)$. From the above inequalities we obtain

$$I_2 \le (\exp(-t\varepsilon E\psi^2(X)/2 + 2t^2A^4))^n = \exp\left(-\frac{\varepsilon^2 n(E\psi^2(X))^2}{64A^4}\right).$$

Next we estimate I_1. Clearly,

$$\begin{aligned} I_1 \le\ & P\left(\sum_{i=1}^n \psi(X_i)I\{|\psi(X_i)| \le z_n\} \ge (1-3\varepsilon)x_n\sqrt{nE\psi^2(X)}\right) \\ & + P\left(\sup_{|s-1|\le\delta}\sum_{i=1}^n(\psi(sX_i) - \psi(X_i))I\{|\psi(X_i)| \le z_n\} \ge \varepsilon x_n\sqrt{nE\psi^2(X)}\right) \\ :=\ & I_{1,1} + I_{1,2}. \end{aligned}$$

Using the elementary inequality $e^x \le 1 + x + x^2/2 + |x|^3e^x$, we have

$$\begin{aligned} I_{1,1} & \le e^{-(1-3\epsilon)x_n\sqrt{n}/z_n} E\exp\left(\frac{1}{z_n\sqrt{E\psi^2(X)}}\sum_{i=1}^n \psi(X_i)I\{|\psi(X_i)| \le z_n\}\right) \\ & = e^{-(1-3\epsilon)x_n^2}\left(E\exp\left(\frac{1}{z_n\sqrt{E\psi^2(X)}}\psi(X)I\{|\psi(X)| \le z_n\}\right)\right)^n \\ & \le e^{-(1-3\epsilon)x_n^2}\left(1 - \frac{1}{z_n\sqrt{E\psi^2(X)}}E\psi(X)I\{|\psi(X)| > z_n\}\right. \\ & \qquad + \frac{1}{2z_n^2E\psi^2(X)}E\psi^2(X)I\{|\psi(X)| \le z_n\} \\ & \qquad \left. + \frac{1}{z_n^3(E\psi^2(X))^{3/2}}E|\psi(X)|^3I\{|\psi(X)| \le z_n\}e^{1/\sqrt{E\psi^2(X)}}\right)^n \\ & \le e^{-(1-3\epsilon)x_n^2}\left(1 + \frac{1}{2z_n^2} + o(1/z_n^2)\right)^n \\ & \le e^{-(1-3\epsilon)x_n^2 + n/(2z_n^2) + o(n/z_n^2)} \le e^{-(1-3\epsilon)^2x_n^2/2} \end{aligned}$$

for sufficiently large n.

Finally, we take on $I_{1,2}$. Let

$$\begin{aligned} \eta_i(s) &= (\psi((1+s\delta)X_i) - \psi(X_i))I\{|\psi(X_i)| \le z_n\}, \qquad -1 \le s \le 1, \\ \eta(s) &= (\psi((1+s\delta)X) - \psi(X))I\{|\psi(X)| \le z_n\}, \qquad -1 \le s \le 1, \\ \Lambda(s) &= \sum_{i=1}^n \eta_i(s) \qquad \text{and} \qquad \Delta = n^{-4/\alpha}. \end{aligned}$$

Without loss of generality, assume $0 < \alpha \le 1$. We have

$$I_{1,2} \le P\left(\sup_{|s|\le 1}\sum_{i=1}^n \eta_i(s) \ge 4\varepsilon^2 x_n\sqrt{n}\right)$$

$$\begin{aligned}
&\le P\left(\max_{-1/\Delta\le k\le 1/\Delta}\Lambda(k\Delta)\ge 2\varepsilon^2 x_n\sqrt{n}\right)\\
&\quad + P\left(\max_{-1/\Delta\le k\le 1/\Delta}\sup_{k\Delta\le s\le (k+1)\Delta}\Lambda(s)-\Lambda(k\Delta)\ge 2\varepsilon^2 x_n\sqrt{n}\right)\\
&\le \sum_{-1/\Delta\le k\le 1/\Delta} P(\Lambda(k\Delta)\ge 2\varepsilon^2 x_n\sqrt{n})\\
&\quad + \sum_{-1/\Delta\le k\le 1/\Delta} P\left(\sum_{i=1}^{n}\sup_{k\Delta\le s\le (k+1)\Delta}|\eta_i(s)-\eta(k\Delta)|\ge 2\varepsilon^2 x_n\sqrt{n}\right)\\
&:= I_{1,3}+I_{1,4}.
\end{aligned}$$

By (R5) and (R6),

$$\begin{aligned}
|E\Lambda(k\Delta)| &\le nE|\psi((1+k\Delta)X)-\psi(X)|I\{|\psi(X)|>z_n\}\\
&\le nKE(1+|\psi(X)|)I\{|\psi(X)|>z_n\}\\
&\le 2nKE|\psi(X)|I\{|\psi(X)|>z_n\}\\
&\le 2nKz_n^{-1}E|\psi(X)|^2 I\{|\psi(X)|>z_n\}\\
&= o(x_n\sqrt{n}).
\end{aligned}$$

Write $\hat{\eta_k}=\eta(k\Delta)-E\eta(k\Delta)$. From the well-known inequality

$$e^x\le 1+x+x^2e^{|x|}, \tag{4.4}$$

it follows that

$$\begin{aligned}
E\exp\left(\frac{8x_n\hat{\eta_k}}{\alpha\varepsilon^2\sqrt{n}}\right) &\le 1+\left(\frac{8x_n}{\alpha\varepsilon^2\sqrt{n}}\right)^2 E\hat{\eta_k}^2\exp\left(\frac{8x_n\hat{\eta_k}}{\alpha\varepsilon^2\sqrt{n}}\right)\\
&\le 1+\left(\frac{8x_n}{\alpha\varepsilon^2\sqrt{n}}\right)^2 E\hat{\eta_k}^2\exp\left(\frac{64x_nKz_n}{\alpha\varepsilon^2\sqrt{n}}\right)\\
&\le 1+\left(\frac{8x_n}{\alpha\varepsilon^2\sqrt{n}}\right)^2 4K(2\delta s_0)^\alpha\exp\left(\frac{64K}{\alpha\varepsilon^2}\right)\\
&\le 1+x_n^2/(\alpha n)\le e^{x_n^2/(\alpha n)}
\end{aligned}$$

as long as δ is chosen to be sufficiently small. Hence

$$\begin{aligned}
P(\Lambda(k\Delta)\ge\varepsilon^2x_n\sqrt{n}) &\le P(\Lambda(k\Delta)-E\Lambda(k\Delta)\ge\varepsilon^2x_n\sqrt{n})\\
&\le e^{-8x_n^2/\alpha}E\exp\left(\frac{8x_n(\Lambda(k\Delta)-E\Lambda(k\Delta))}{\alpha\varepsilon^2\sqrt{n}}\right)\\
&= e^{-8x_n^2/\alpha}\left(E\exp\left(\frac{8x_n\hat{\eta_k}}{\alpha\varepsilon^2\sqrt{n}}\right)\right)^n\\
&< e^{-7x_n^2/\alpha}.
\end{aligned}$$

Thus, for $x > \sqrt{\log n}$,

$$I_{1,3} \le (2/\Delta)e^{-7x_n^2/\alpha} \le e^{-2x_n^2}. \tag{4.5}$$

Turning to $I_{1,4}$, we note by (R6),

$$\begin{aligned}\sum_{i=1}^{n} E \sup_{k\Delta \le s \le (k+1)\Delta} &|\eta_i(s) - \eta(k\Delta)| \\ &\le n(K\Delta^\alpha)^{1/2} \le \sqrt{K} = o(x_n\sqrt{n}).\end{aligned}$$

Similar to estimating $I_{1,3}$, we obtain

$$\begin{aligned}I_{1,4} &\le P\Bigg(\sum_{i=1}^{n} \sup_{k\Delta \le s \le (k+1)\Delta} |\eta_i(s) - \eta(k\Delta)| \\ &\qquad - E \sup_{k\Delta \le s \le (k+1)\Delta} |\eta_i(s) - \eta(k\Delta)| \ge \varepsilon^2 x_n \sqrt{n}\Bigg) \\ &\le e^{-2x_n^2}.\end{aligned}$$

Putting things together, we arrive at

$$\limsup_{n\to\infty} x_n^{-2} \log P(T_n^s \ge x_n) \le -\frac{(1-3\varepsilon)^3}{2}.$$

This proves (4.1) by the arbitrariness of ε.

We now turn to (4.2). For any $0 < \varepsilon < 1/4$, take $0 < \delta = \delta(\varepsilon) < \varepsilon$ to be small enough so that (4.3) is satisfied. Notice that

$$\begin{aligned}P(T_n^s \ge x_n) &\ge P(T_n^s \ge x_n, |s_n - 1| \le \delta/2) \\ &\ge P\left(\inf_{|s-1|\le\delta/2} \frac{\sum_{i=1}^n \psi(X_i/s)}{\sqrt{\sum_{i=1}^n \psi^2(X_i/s)}} \ge x_n, |s_n - 1| \le \delta/2\right) \\ &\ge P\left(\inf_{|s-1|\le\delta} \frac{\sum_{i=1}^n \psi(sX_i)}{\sqrt{\sum_{i=1}^n \psi^2(sX_i)}} \ge x_n\right) - P(|s_n - 1| \ge \delta/2) \\ &\ge P\left(\inf_{|s-1|\le\delta} \frac{\sum_{i=1}^n \psi(sX_i)}{\sqrt{\sum_{i=1}^n \psi^2(sX_i)}} \ge x_n, \max_{i\le n}|\psi(X_i)| \le z_n\right) - \eta^n \\ &:= J_1 - \eta^n\end{aligned}$$

and

$$\begin{aligned}J_1 \ge\ & P\left(\inf_{|s-1|\le\delta} \frac{\sum_{i=1}^n \psi(sX_i)}{\sqrt{nE\psi^2(X)}} \ge (1+\varepsilon)x_n, \max_{i\le n}|\psi(X_i)| \le z_n\right) \\ &- P\left(\sum_{i=1}^{n} \sup_{|s-1|\le\delta} \psi^2(sX_i) \ge (1+\varepsilon)^2 nE\psi^2(X), \max_{i\le n}|\psi(X_i)| \le z_n\right)\end{aligned}$$

$$
\begin{aligned}
&\geq P\left(\sum_{i=1}^{n}\psi(X_i) \geq (1+2\varepsilon)x_n\sqrt{nE\psi^2(X)}, \max_{i\leq n}|\psi(X_i)| \leq z_n\right)\\
&\quad - P\left(\sup_{|s-1|\leq\delta}\sum_{i=1}^{n}(\psi(sX_i)-\psi(X_i)) \geq \varepsilon x_n\sqrt{nE\psi^2(X)}, \max_{i\leq n}|\psi(X_i)| \leq z_n\right)\\
&\quad - P\left(\sum_{i=1}^{n}\sup_{|s-1|\leq\delta}\psi^2(sX_i) \geq (1+\varepsilon)^2 nE\psi^2(X), \max_{i\leq n}|\psi(X_i)| \leq z_n\right)\\
&:= J_{1,1} - J_{1,2} - J_{1,3}.
\end{aligned}
$$

Let

$$
\begin{aligned}
\zeta_i &= \sup_{|s-1|\leq\delta}\psi^2(sX_i)I\{|\psi(X_i)|\leq z_n\} \qquad \text{and}\\
\zeta &= \sup_{|s-1|\leq\delta}\psi^2(sX)I\{|\psi(X)|\leq z_n\}.
\end{aligned}
$$

From (4.4) and (R5) we obtain

$$
\begin{aligned}
&E\exp\left(\frac{4x_n^2(\zeta-E\zeta)}{\varepsilon nE\psi^2(X)}\right)\\
&\quad\leq 1+\left(\frac{4x_n^2}{\varepsilon nE\psi^2(X)}\right)^2 E(\zeta-E\zeta)^2\exp\left(\frac{4x_n^2|\zeta-E\zeta|}{\varepsilon nE\psi^2(X)}\right)\\
&\quad\leq 1+\left(\frac{4x_n^2}{\varepsilon nE\psi^2(X)}\right)^2 \exp\left(\frac{8K^2}{\varepsilon E\psi^2(X)}\right)\\
&\qquad\cdot 4K^2E(1+\psi^4(X))I\{|\psi(X)|\leq z_n\}\\
&\quad\leq 1+\left(\frac{4x_n^2}{\varepsilon nE\psi^2(X)}\right)^2 \exp\left(\frac{8K^2}{\varepsilon E\psi^2(X)}\right)4K^2z_n^2o(1)\\
&\quad\leq 1+x_n^2/n\leq e^{x_n^2/n},
\end{aligned}
$$

provided that n is sufficiently large. Therefore,

$$
\begin{aligned}
J_{1,3} &\leq P\left(\sum_{i=1}^{n}\zeta_i \geq (1+\varepsilon)^2 nE\psi^2(X)\right)\\
&\leq P\left(\sum_{i=1}^{n}(\zeta_i-E\zeta_i) \geq \varepsilon nE\psi^2(X)\right)\\
&\leq e^{-4x_n^2}E\exp\left(\frac{4x_n^2\sum_{i=1}^{n}(\zeta_i-E\zeta_i)}{\varepsilon nE\psi^2(X)}\right)\\
&= e^{-4x_n^2}\left(E\exp\left(\frac{4x_n^2(\zeta-E\zeta)}{\varepsilon nE\psi^2(X)}\right)\right)^n \leq e^{-3x_n^2}
\end{aligned}
$$

for sufficiently large n.

In terms of (4.5), we have

$$J_{1,2} \le P\left(\sup_{|s-1|\le\delta}\sum_{i=1}^{n}(\psi(sX_i)-\psi(X_i))I\{|\psi(X_i)|\le z_n\}\ge \varepsilon x_n\sqrt{nE\psi^2(X)}\right)$$
$$\le e^{-2x_n^2}.$$

To estimate $J_{1,1}$, let $\tau, \tau_1, \tau_2, \ldots$ be i.i.d. random variables with the distribution $G(\cdot)$ of $\psi(X)$ conditional on $|\psi(X)| \le z_n$. Then

$$dG(x) = \frac{I\{|x|\le z_n\}}{P(|\psi(X)|\le z_n)}dP(\psi(X)\le x). \tag{4.6}$$

Using Lemma 5.1 of Griffin and Kuelbs (1989), we obtain

$$J_{1,1} = P\left(\sum_{i=1}^{n}\tau_i \ge (1+2\varepsilon)x_n\sqrt{nE\psi^2(X)}\right)P(\max_{i\le n}|\psi(X_i)|\le z_n).$$

It is easy to see that there exists $w_n = o(1)$ such that

$$\begin{aligned}P(\max_{i\le n}|\psi(X_i)|\le z_n) &= (1-P(|\psi(X)|>z_n))^n\\ &\ge (1-z_n^{-2}w_n))^n \ge e^{-x_n^2 w_n}.\end{aligned}$$

Let $h = (1+3\varepsilon)x_n/\sqrt{nE\psi^2(X)}$. Clearly

$$\begin{aligned}P(|\psi(X)|\le z_n\}Ee^{h\tau} &= Ee^{h\psi(X)}I\{|\psi(X)|\le z_n\}\\ &= E\left(1+h\psi(X)+\frac{h^2\psi^2(X)}{2}+O(1)h^3|\psi(X)|^3e^{h|\psi(X)|}I\{|\psi(X)|\le z_n\}\right)\\ &= P(|\psi(X)|\le z_n) - hE\psi(X)I\{|\psi(X)|\le z_n\}\\ &\quad+\frac{h^2E\psi^2(X)}{2}+O(1)h^3E|\psi^3(X)|I\{|\psi(X)|\le z_n\}\\ &= 1+\frac{h^2E\psi^2(X)}{2}+o(1)h^2+o(z_n^{-2})\\ &= 1+\frac{h^2E\psi^2(X)}{2}+o(h^2).\end{aligned}$$

Noting that $P(|\psi(X)|\le z_n) = 1-o(z_n^{-2})$, we get

$$Ee^{h\tau} = 1+\frac{h^2E\psi^2(X)}{2}+o(h^2) = 1+\frac{(1+3\varepsilon)^2x_n^2}{2n}+o(x_n^2/n).$$

Similarly, we have

$$E\tau e^{h\tau} = hE\psi^2(X)+o(h) = \frac{(1+3\varepsilon)x_n\sqrt{E\psi^2(X)}}{\sqrt{n}}+o(x_n/\sqrt{n}),$$

and

$$E\tau^2 e^{h\tau} = E\psi^2(X) + o(1).$$

Therefore

$$\begin{aligned}
m_h &:= \frac{E\tau e^{h\tau}}{Ee^{h\tau}} \\
&= \frac{(1+3\varepsilon)x_n\sqrt{E\psi^2(X)}/\sqrt{n} + o(x_n/\sqrt{n})}{1+\frac{(1+3\varepsilon)^2 x_n^2}{2n} + o(x_n^2/n)} \\
&= \frac{(1+3\varepsilon)x_n\sqrt{E\psi^2(X)}}{\sqrt{n}} + o(x_n/\sqrt{n}) \\
&= \frac{(1+2\varepsilon)x_n\sqrt{E\psi^2(X)}}{\sqrt{n}} + \frac{\varepsilon x_n\sqrt{E\psi^2(X)}}{\sqrt{n}} o(x_n/\sqrt{n}) \\
&\geq \frac{x_n\sqrt{E\psi^2(X)}}{\sqrt{n}} + \frac{2\sqrt{E\tau^2 e^{h\tau}}}{\sqrt{n}}.
\end{aligned}$$

Consequently, applying Lemma 3.1 of Shao (1994) yields

$$\begin{aligned}
J_{1,1} &= P\left(\sum_{i=1}^{n}\tau_i \geq \frac{(1+2\varepsilon)x_n\sqrt{E\psi^2(X)}}{\sqrt{n}}n\right) \\
&\geq 0.75(Ee^{h\tau})^n \exp(-nhm_h - 4h\sqrt{n}E\psi^2(X)) \\
&\geq 0.75\exp\left(-\frac{(1+3\varepsilon)^2 x_n^2}{2} + o(x_n^2)\right).
\end{aligned}$$

Putting things together, we conclude that

$$\liminf_{n\to\infty} x_n^{-2}\log P(T_n^s \geq x_n) \geq -\frac{(1+3\varepsilon)^2}{2}.$$

From the arbitrariness of ε follows (4.2), and therefore (2.7). It is easy to see, along the same proof here, that (2.8) holds.

Acknowledgements

The research is partly supported by a grant from the University of Illinois Campus Research Board and a National University of Singapore's Research Project. We thank three referees for their helpful suggestions.

REFERENCES

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. Statist.*, **31**, 276–295.

Bahadur, R. R. (1965). An optimal property of the likelihood ratio statistic, *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, **1**, 13–26.

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

Efron, B. (1969). Student's t-test under symmetry conditions, *J. Amer. Statist. Assoc.*, **64**, 1278–1302.

Griffin, P. S. and Kuelbs, J. (1989). Self-normalized laws of the iterated logarithm, *Ann. Probab.*, **17**, 1571–1601.

Groeneboom, P. and Oosterhoff, J. (1977). Bahadur efficiency and probability of large deviations, *Statist. Neerlandica*, **31**, 1–24.

Hampel, F. R. (1974). The influence and its role in robust estimation, *J. Amer. Statist. Assoc.*, **69**, 383–393.

Hampel, F. R., Ronchetti, E., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

He, X., Simpson, D. G. and Portnoy, S. L. (1990). Breakdown robustness of tests, *J. Amer. Statist. Assoc.*, **85**, 446–452.

Huber, P. J. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, **35**, 73–101.

Huber, P. J. (1965). A robust version of the probability ratio test, *Ann. Math. Statist.*, **36**, 1753–1758.

Kallenberg, W. C. M. and Ledwina, T. (1987). On local and nonlocal measures of efficiency, *Ann. Statist.*, **15**, 1401–1420.

Lambert, D. (1981). Influence functions for tests, *J. Amer. Statist. Assoc.*, **76**, 649–657.

Lambert. D. and Hall, W. J. (1982). Asymptotic lognormality of P-values, *Ann. Statist.*, **10**, 44–64.

Lloyd, C. (1994). Approximate pivots from M-estimators, *Statistica Sinica*, **4**, 701–714.

Rukhin, A. L. (1993). On the Bahadur efficiency of a t-test, *Sankhyā Ser. A*, **55**, 159–163.

Shao, Q. M. (1994). Self-normalized Large Deviations, *Ann. Probab.* (to appear in 1997).