Comparison of Optimal Location Estimators

by

Friedrich-Wilhelm Scholz
Grad. (University of Göttingen) 1965

Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Statistics

in the

Graduate Division

of the

University of California, Berkeley

Approved:

. . . .
Leonard A. Marascuilo

. . . .
R.J. Beran

. . . .
E.L. Lehmann

Table of Contents

Comparison of Optimal Location Estimators*

Ph.D. Dissertation, December, 1971

by Friedrich-Wilhelm Scholz

Department of Statistics

University of California, Berkeley

Abstract

This study is concerned with the problem of estimating the center of a symmetric distribution. Three classes of estimators have been considered in the literature: linear combinations of order statistics, estimators derived from rank tests and maximum likelihood type estimators, which are referred to as L-, R- and M^*-estimators respectively. Under regularity conditions these estimators are known to be asymptotically normal, and their performance is judged by their asymptotic variances. From each of the above classes we select estimators $L(F)$, $R(F)$ and $M^*(F)$, which are optimal for a sample arising from a given underlying distribution F (satisfying certain regularity conditions). They are optimal in the sense that their asymptotic variances cannot be improved upon by any other location invariant estimator when the underlying distribution is F. The principal aim of the present investigation is to compare these three estimators for underlying distributions $H \neq F$. It is shown that the asymptotic variance of $R(F)$ is always less or equal to the asymptotic variance of $L(F)$ for all underlying distributions H, satisfying certain regularity conditions. Under more stringent conditions on F, it is shown that a similar statement holds in the comparison of $R(F)$ and $M^*(F)$, where again $R(F)$ is the superior estimator.

For the special case where F is the normal distribution, a result of this kind was given by Chernoff and Savage in the context of testing. They showed that the asymptotic relative efficiency of the normal scores rank test relative to the t-test never falls below one. Mikulski then showed that this result is specific to the normal distribution. For a given underlying distribution F he constructs a best parametric test and a best nonparametric test and compares them for other underlying distributions. Mikulski's result depends strongly on the fact that his best parametric test does not

*This research was partially supported by National Science Foundation Grant GP-8690 and by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant AF-AFOSR-1312-67

possess certain invariance properties. We construct best parametric tests which have these invariance properties and obtain results similar to those of Chernoff and Savage.

Acknowledgment

I would like to express my deep gratitude to Professor Erich L. Lehmann for his advice and guidance throughout the course of this research.

Comparison of Optimal Location Estimators*

Ph.D. Dissertation, December, 1971

by Friedrich-Wilhelm Scholz

Department of Statistics

University of California, Berkeley

0. Introduction and Summary

The problem of estimating the center of symmetry of a distribution has been treated quite extensively in the statistical literature. Various classes of estimators have found a special interest and we will deal with three of them: linear combinations of order statistics, estimators that are derived from rank tests, and maximum likelihood type estimators, as outlined in Section 1.1. The performance of a given estimator depends strongly on the underlying distribution of the given sample. Since it is difficult to study the behavior of estimators for finite sample sizes, most research has focussed on the asymptotic behavior of these estimators. It is hoped that the asymptotic results provide useful approximations to the finite sample size case. Most of the estimators commonly studied are, under suitable regularity conditions, normally distributed around the parameter to be estimated, with asymptotic variances depending on the underlying distribution. We therefore have a simple criterion, the asymptotic variance, for comparing the performance of different estimators. The usual approach is then the following. One considers two interesting estimators, say for example the sample mean and the sample median, and compares their asymptotic variances for various underlying distributions. Typically one finds that neither one of them is uniformly better than the other. This will happen in particular if both estimators are in some sense optimal at different underlying distributions, as is the case with the sample mean and sample median, which are optimal for the normal and double exponential

*This research was partially supported by National Science Foundation Grant GP-8690 and by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Grant AF-AFOSR-1312-67

distribution, respectively. Thus in comparing two such estimators a more differentiated evaluation must be made and one must weigh the advantages and disadvantages of the estimators.

It may of course happen that the performance of one estimator is never worse than the performance of the other. Chernoff and Savage (1958) give one example of such a comparison, even though they consider this problem in terms of testing. For the problem of testing for shift in two samples, they showed that the asymptotic relative efficiency of the normal scores test relative to the t-test never falls below one. Using the results by Hodges and Lehmann (1963), one can rephrase this result as follows for the estimation problem: The estimator derived from the normal scores rank test has an asymptotic variance which is always less than or equal to the asymptotic variance of the sample mean. Equality of the asymptotic variances occurs if and only if the underlying distribution is normal, in which case both estimators are optimal in some sense. This result and our previous remarks suggest that we take the following approach in finding other examples of this kind.

We will compare only estimators with each other which are optimal for a fixed given underlying distribution F. By 'optimal' we mean that the asymptotic variance of these estimators cannot be improved upon by any other location invariant estimator. In Section 1.2 we give the construction of three estimators which are optimal for F. These are denoted by $L(F)$, $R(F)$ and $M^*(F)$ and they are respectively a particular linear combination of order statistics, the estimator derived from a particular rank test and a particular maximum likelihood type estimator. In Chapter 2 we examine the asymptotic variances of these three estimators when the underlying distribution is different from F.

Section 2.1 deals with the comparison of $L(F)$ and $R(F)$ and it is seen that the result of Chernoff and Savage (as it was rephrased for the estimation problem) is a special case of Theorem 1, which roughly states the following: If F is sufficiently regular, then the asymptotic variance of $R(F)$ is always less or equal to the asymptotic variance of $L(F)$ for all underlying distributions H satisfying certain regularity conditions. Section 2.2 deals with the comparison of $R(F)$ and $M^*(F)$ and an analogous result is obtained in Theorem 2. Again $R(F)$ turns out to be the superior one of the two estimators. In Theorem 2 we impose a certain concavity condition on F. This condition, a counterexample to Theorem 2 when certain regularity conditions are not met, and the comparison of $M^*(F)$ and $L(F)$ are discussed in Section 2.3 and in the Appendix.

In Chapter 3 we discuss a paper by Mikulski (1963). The main theorem of this paper states that the result arrived at by Chernoff and Savage in the context of hypothesis testing is specific to the normal distribution. Mikulski considers the two-sample shift problem and constructs parametric and nonparametric tests which are optimal for a given underlying distribution F. Then he studies the asymptotic relative efficiency of the nonparametric test relative to the parametric test for other underlying distributions H and finds that this efficiency can fall below one if F is not the normal distribution. His method of proving this depends strongly on the fact that his parametric test is not location and scale invariant. Since the rank tests and the t-test have this invariance property, we propose a location and scale invariant parametric test and obtain results similar to the one given by Chernoff and Savage.

1. Three Types of Location Estimators.

1.1 Definitions.

Let $X_1, \ldots, X_n$ be independent identically distributed random variables with distribution $H_\mu(x) = H(x-\mu)$. We assume that H has a density $h(x)$ and is symmetric around zero., i.e., $H(x) = 1 - H(-x)$, but that otherwise H is unknown. We want to estimate the unknown location parameter μ, and the performance of various estimators will be judged by their asymptotic variances. Under general regularity conditions, each of the considered estimators is asymptotically normal with mean μ and asymptotic variances given below.

We introduce the following notation:

$$\boldsymbol{X}_n = (X_1, \ldots, X_n) \qquad \text{and} \qquad b\boldsymbol{X}_n + a = (bX_1 + a, \ldots, bX_n + a)$$

for real numbers a and b. An estimator $T_n(\boldsymbol{X}_n)$ is location invariant if $T_n(\boldsymbol{X}_n + a) = T_n(\boldsymbol{X}_n) + a$ for all real numbers a. An estimator $T_n(\boldsymbol{X}_n)$ is scale invariant if $T_n(b\boldsymbol{X}_n) = bT_n(\boldsymbol{X}_n)$ for all real numbers b. The dependence of $T_n(\boldsymbol{X}_n)$ on the sample $\boldsymbol{X}_n$ is usually understood and we shall simply write T_n whenever no confusion will arise thereby. Since all estimators to be considered here will be translation invariant, we shall study their asymptotic distribution without loss of generality in the case of $\mu = 0$.

One popular class of estimators is the class of linear combinations of order statistics, which we shall call L-estimators. Let $X_{(1)} \leq \ldots \leq X_{(n)}$ be the order statistics of the sample and let g be a function mapping the open interval $(0,1)$ into the set of real numbers, $\mathbb{R}$, such that $g(t) = g(1-t)$ and $\int_0^1 g(t)dt = 1$. We define the L-estimator corresponding to g by

$$L_n = L_n(g) = \frac{1}{n}\sum_{i=1}^{n} g\left(\frac{i}{n+1}\right) X_{(i)} \ .$$

Under general regularity conditions on g and H we have

$$\sqrt{n}\ L_n(g) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma_g^2(H)\right) \ ,$$

where

$$\sigma_g^2(H) = \int_0^1 U^2(t)\ dt \qquad \text{with} \qquad U(t) = \int_{1/2}^{t} \left[\frac{g(u)}{h(H^{-1}(u))}\right]\ du \ .$$

(cf. Chernoff, Gastwirth and Johns (1967)). Denote the class of distributions H for which this asymptotic normality holds by $CL(g)$. We observe that L-estimators are location and scale invariant.

Another class of estimators was introduced by Hodges and Lehmann (1963). These estimators, called R-estimators, are derived from rank tests or, more precisely, from the rank statistics which are employed in these rank tests. Let J be a function mapping the open interval $(0,1)$ into the set of real numbers such that $J(t) = -J(1-t)$. For $r \in \mathbb{R}$ denote by $S_i(r)$ the rank of $|X_i - r|$ among $|X_1 - r|$, ..., $|X_n - r|$. We then define the following linear rank statistic:

$$T_n(r) = \sum_{i=1}^{n} J\left(\frac{S_i(r)}{n+1}\right) I_i(r) \ ,$$

where

$$I_i(r) = \begin{cases} 1 & \text{if} \quad X_i > r \\ 0 & \text{if} \quad X_i \le r \end{cases} \ .$$

Let

$$r_n^* = \sup\{r : T_n(r) > 0\} \qquad \text{and} \qquad r_n^{**} = \inf\{r : T_n(r) < 0\} \ .$$

The R-estimator corresponding to J is then defined by

$$R_n = R_n(J) = \frac{r_n^* + r_n^{**}}{2} \ .$$

Under regularity conditions on J and H we have

$$\sqrt{n}\, R_n(J) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma_J^2(H)\right),$$

where

$$\sigma_J^2(H) = \left(\int_0^1 J^2(u)\, du\right) \left(\int \frac{d}{dx}[J(H(x))]\, dH(x)\right)^{-2}$$

(cf. Hodges and Lehmann (1963) and Puri and Sen (1971)). Denote the class of distributions H for which this asymptotic normality holds by $CR(J)$. Again we observe that R-estimators are location and scale invariant.

A third type of estimator was studied by Huber (1964). Let $\psi : \mathbb{R} \longrightarrow \mathbb{R}$ with $\psi(-x) = -\psi(x)$, let ψ be nondecreasing and let $M_n(\psi)$ be defined to be the solution M of

$$\sum_{i=1}^{n} \psi(X_i - M) = 0\ .$$

For particular choices of the function ψ the estimators $M_n(\psi)$, called M-estimators, are identical with maximum likelihood estimators of the location parameter of some distribution. Under general regularity conditions on ψ and H (see Huber, 1967), we have

$$\sqrt{n}\, M_n(\psi) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma_\psi^2(H)\right)$$

where

$$\sigma_\psi^2(H) = \int \psi^2(x)\, dH(x) \left(\int \psi'(x)\, dH(x)\right)^{-2}\ .$$

The M-estimators are location invariant, but in general not scale invariant. This is not desirable, since with the choice of the ψ-function we also express some judgment on the scale of the underlying distribution H. Since we do not assume any knowledge about the scale of H, it seems advisable to construct a scale invariant version of the M-estimator. Huber (1964) suggests a way of achieving this in the case of a special ψ-function (Huber's proposal two) and for more general ψ-functions in his 1970 paper. We shall assume that ψ is continuously differentiable with $\psi'(x) > 0$ for all $x \in \mathbb{R}$ and that $\psi(x) = -\psi(-x)$.

Definition 1. Let $M_n^*(\psi)$ and $S_n^*(\psi)$ be the unique solutions M and S of the following system of equations:

$$(0) \qquad \frac{1}{n}\sum_{i=1}^{n} \psi\left(\frac{X_i - M}{S}\right) = 0 \qquad \text{and} \qquad \frac{1}{n}\sum_{i=1}^{n} \psi^2\left(\frac{X_i - M}{S}\right) = \beta$$

where β is a fixed number satisfying: $0 < \beta < \sup\{\psi^2(x) : x \in \mathbb{R}\}$.

In order to justify this definition, we must show that the solutions M and S exist and are unique. This is done in Appendix A. Huber's paper (1967) establishes the consistency and asymptotic normality of the estimates $(M_n^*(\psi), S_n^*(\psi))$ under regularity conditions on ψ and H. In particular we have

$$\sqrt{n}\, M_n^*(\psi) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, {\sigma_\psi^*}^2(H)\right)$$

where

$${\sigma_\psi^*}^2(H) = \frac{\int \psi^2(x/\tau_H)\, dH(x)}{\left(\int \psi'(x/\tau_H)\, dH(x)\right)^2}\ \tau_H^2$$

with τ_H determined by

$$\int \psi^2(x/\tau_H)\, dH(x) = \beta\ .$$

We denote by $CM^*(\psi)$ the class of distributions H for which this asymptotic normality for $M_n^*(\psi)$ is valid. From the definition it is clear that $M_n^*(\psi)$ is location and scale invariant. We will call these estimators M^*-estimators.

1.2 Optimal Estimators for *F*.

In the previous section we introduced three classes of location and scale invariant estimators for the center of symmetry of an unknown distribution H. Now we will assume that the sample arises from a distribution $F((x-\mu)/\sigma)$, where F is completely known, symmetric around zero and satisfies certain regularity conditions. The parameters μ and σ are unknown. We will construct L-, R- and M^*-estimators for μ which are optimal at F in the following sense: among all location invariant estimators for μ, they have the smallest possible asymptotic variance, whatever μ and σ may be. The optimality and general correspondence between these estimators was shown by Jaeckel (1971). We denote these three optimal estimators by $L(F)$, $R(F)$ and $M^*(F)$. Since they are location and scale invariant, we may restrict ourselves without loss of generality to the case $\mu = 0$ and $\sigma = 1$.

We will impose the following regularity conditions on F:

i) $F(x)$ has a density $f(x)$ with the properties $f(x) = f(-x)$ on $\mathbb{R}$ and $f(x) > 0$ on $\mathbb{R}$;

ii) f is continuously differentiable on $\mathbb{R}$;

iii) $I(f) = \int \psi_f^2(x) f(x)\, dx < \infty, \quad \text{where} \quad \psi_f(x) = -f'(x)/f(x).$

The class of distributions F satisfying i)–iii) will be denoted by $\mathcal{F}$.

Definition of $M^*(F)$: Let $\psi(x) = \psi_f(x)$ and set $\beta = \int \psi^2(x)\, dF(x)$. Assume that $\psi(x)$ has a continuous derivative $\psi'(x) > 0$ on $\mathbb{R}$ and define: $M^*(F) = M_n^\star(\psi)$. If $F \in CM^*(\psi)$, then $\sqrt{n}\ M^*(F)$ is asymptotically normal with mean zero and asymptotic variance

$$\sigma_F^2(M^*(F)) = {\sigma_\psi^*}^2(F) = \left(\int \psi^2(x/\tau_F)\, dF(x)\right)\left(\int \psi'(x/\tau_F)\, dF(x)\right)^{-2} \tau_F^2 \ ,$$

where τ_F satisfies

$$\int \psi^2(x/\tau_F)\, dF(x) = \beta = \int \psi^2(x)\, dF(x) \ ,$$

i.e., $\tau_F = 1$. Hence

$$\sigma_F^2(M^*(F)) = I(f)\left(\int \psi'(x)\, dF(x)\right)^{-2} = [I(f)]^{-1} \ .$$

It follows from results of LeCam (1953) and Hajek (1971) that $[I(f)]^{-1}$ is the smallest possible asymptotic variance that can be achieved by any asymptotically normal location invariant estimator. Thus $M^*(F)$ is optimal at F.

Definition of $L(F)$: Assume that $\psi_F(x) = \psi(x)$ has a continuous derivative ψ' except at a finite number of points. Let $g(t) = [I(f)]^{-1}\psi'(F^{-1}(t))$ and define $L(F) = L_n(g)$. If $F \in CL(g)$, then $\sqrt{n}\ L(F)$ is asymptotically normal with mean zero and asymptotic variance

$$\sigma_F^2(L(F)) = \sigma_g^2(F) = \int_0^1 U^2(t)\, dt \ ,$$

where

$$U(t) = \int_{\frac{1}{2}}^t \frac{\psi'(F^{-1}(u))}{f(F^{-1}(u))}\, du \ \ [I(f)]^{-1} = \psi(F^{-1}(t))[I(f)]^{-1},$$

hence

$$\sigma_F^2(L(F)) = \int_0^1 \psi^2(F^{-1}(t))\, dt \ \ [I(f)]^{-2} = [I(f)]^{-1} \ .$$

Thus $L(F)$ is also optimal at F.

Definition of $R(F)$: Assume that $\psi_f(x) = \psi(x)$ has a continuous derivative ψ' except at a finite number of points. Let $J(t) = \psi(F^{-1}(t))$ and define $R(F) = R_n(J)$. If $F \in CR(J)$, then $\sqrt{n}\, R(F)$ is asymptotically normal with mean zero and asymptotic variance

$$\sigma_F^2(R(F)) = \sigma_J^2(F) = \left(\int_0^1 \psi^2(F^{-1}(t))\, dt \right) \left(\int \psi'(x)\, dF(x) \right)^{-2} = [I(f)]^{-1}$$

which implies that $R(F)$ is optimal at F.

2. Comparison of the Three Optimal Estimators for F

We will now investigate the relationship between the asymptotic variances of $L(F)$, $R(F)$ and $M^*(F)$ when the sample comes from a symmetric distribution H different from F. We also assume that H satisfies the various regularity conditions to ensure the asymptotic normality of the estimators.

2.1. Comparison of $R(F)$ and $L(F)$.

We assume that the distribution F which defines $R(F)$ and $L(F)$ satisfies the conditions:

A1: $F \in \mathcal{F}$, where $\mathcal{F}$ is defined on p. 7

A2: ψ_f has a nonnegative continuous derivative except at a finite number of points;

A3: $F \in CR(J) \cap CL(g)$, where

$$J(t) = \psi_f(F^{-1}(t)) \qquad \text{and} \qquad g(t) = [I(f)]^{-1} \psi_f'(F^{-1}(t))\ .$$

On H we will impose the following restrictions:

B1: H has a density $h(x)$ and is symmetric around zero;

B2: $h(x) > 0$ for $x \in \{x : 0 < H(x) < 1\}$;

B3: $H \in CR(J) \cap CL(g)$, where J and g are the same as in A3.

When H is the underlying distribution of the sample, the asymptotic variances of $\sqrt{n}\ R(F)$ and $\sqrt{n}\ L(F)$ are given by:

$$\sigma_H^2(R(F)) = I(f) \left(\int \frac{\psi_f'(F^{-1}(H(x)))}{f(F^{-1}(H(x)))}\ h^2(x)\ dx \right)^{-2}$$

and

$$\sigma_H^2(L(F)) = \int_0^1 U^2(t)\ dt \qquad \text{with} \qquad U(t) = \int_{\frac{1}{2}}^t \frac{\psi_f'(F^{-1}(u))}{h(H^{-1}(u))}\ du\ \ [I(f)]^{-1}\ .$$

In the special case where F is equal to the normal distribution Φ, it was shown by Chernoff and Savage (1958) that $\sigma_H^2(R(\Phi)) \leq \sigma_H^2(L(\Phi))$ and equality holds if and only if $H(x) = \Phi(ax)$ for some $a > 0$. Gastwirth and Wolff (1968) gave a simpler proof for the same result, and we will use their method to prove the following theorem.

Theorem 1. Let F satisfy conditions A1-A3. Then

$$\sigma_H^2(R(F)) \leq \sigma_H^2(L(F)) \tag{1}$$

for all H satisfying B1-B3. If further $\psi_f' > 0$ where it is defined, then equality holds in (1) if and only if $H(x) = F(ax)$ for some $a > 0$.

Proof: Let $\psi = \psi_f$ and observe that $\psi'(F^{-1}(t))[I(f)]^{-1}$ is a density on the interval $[0,1]$, since $\psi' \geq 0$ and

$$\int_0^1 \psi'(F^{-1}(t))\ dt = I(f)\ .$$

Using Jensen's inequality, we obtain

$$\begin{aligned}
\left[\int \frac{\psi'(F^{-1}(H(x)))}{f(F^{-1}(H(x)))}\ h^2(x)\ dx\right]^{-1} &= \left[\int_0^1 \frac{I(f)h(H^{-1}(x))}{f(F^{-1}(t))} \frac{\psi'(F^{-1}(t))}{I(f)}\ dt\right]^{-1} \\
&\leq \int_0^1 \frac{f(F^{-1}(t))}{h(H^{-1}(t))}\ \frac{\psi'(F^{-1}(t))}{[I(f)]^2}\ dt \\
&= [I(f)]^{-2}\ f(F^{-1}(t)) \int_{\frac{1}{2}}^t \frac{\psi'(F^{-1}(u))}{h(H^{-1}(u))}\ du \Bigg|_0^1 \\
&\quad + [I(f)]^{-2} \int_0^1 I(f)U(t)\psi(F^{-1}(t))\ dt = E\ .
\end{aligned}$$

Since

$$\int_{\frac{1}{2}}^{1} U^2(t)\,dt < \infty \quad \Longrightarrow \quad \int_0^{\infty} U(F(x))f(x)\,dx < \infty$$

$$\Longrightarrow U(F(x_n))f(x_n) \longrightarrow 0 \text{ for some sequence } x_n \to \infty$$

$$\Longrightarrow U(t_n)f(F^{-1}(t_n)) \longrightarrow 0 \text{ for some sequence } t_n \longrightarrow 1, \text{ we obtain}$$

$$\begin{aligned} E &= [I(f)]^{-1} \int_0^1 U(t)\psi(F^{-1}(t))\,dt \\ &\leq [I(f)]^{-1} \left(\int_0^1 U^2(t)\,dt\right)^{\frac{1}{2}} \left(\int_0^1 \psi^2(F^{-1}(t))\,dt\right)^{\frac{1}{2}} = \left(\int_0^1 U^2(t)\,dt\ \ [I(f)]^{-1}\right)^{\frac{1}{2}}, \end{aligned}$$

thus

$$\sigma_H^2(R(F)) \leq \sigma_H^2(L(F))\ . \tag{1}$$

If in addition we know that $\psi' > 0$ where it exists then it is seen immediately from the above inequalities that equality in (1) holds if and only if $H(x) = F(ax)$ for some $a > 0$. This concludes the proof.

The last assertion of Theorem 1 is not necessarily true without the condition $\psi' > 0$, as can be seen by the following example. Let

$$\psi_k(x) = x \quad \text{for } |x| \leq k \qquad \text{and} \qquad \psi_k(x) = k\,\text{sign}(x) \quad \text{otherwise} \ \ (k > 0)\ .$$

This ψ_k-function plays an important role in Huber's paper (1964). It corresponds to a distribution F_k which is normal in the middle but has double exponential tails. Since $\psi_k'(x) = 0$ for $|x| > k$, it is seen that the asymptotic variance $\sigma_H^2(R(F_k))$ is not changed if one modifies $H(x)$ for those x for which $H(|x|) > F_k(k)$. A similar remark applies to $\sigma_H^2(L(F_k))$. Thus

$$\sigma_H^2(R(F_k)) = \sigma_H^2(L(F_k)) \quad \text{for all } H \text{ which satisfy } H(x) = F_k(x) \text{ for } |x| \leq k.$$

2.2 Comparison of $R(F)$ and $M^*(F)$.

We assume that the distribution F, which defines $R(F)$ and $M^*(F)$, satisfies the following conditions:

$A'1$: $F \in \mathcal{F}$, where $\mathcal{F}$ is defined on page 7;

$A'2$: $\psi_f(x) = \psi(x) = -f'(x)/f(x)$ is twice continuously differentiable for $x > 0$ and $\psi'(x) > 0$ for $x > 0$;

$A'3$: $F \in CM^*(\psi) \cap CR(J)$, where $J(t) = J_f(t) = \psi(F^{-1}(t))$;

$A'4$: $1/J'(t))$ is concave on the interval $(\frac{1}{2}, 1)$.

On the distribution H of the underlying sample we will impose the following restrictions:

$B'1$: H has a density $h(x)$ and is symmetric around zero;

$B'2$: $H \in CR(J_f) \cap CM^*(\psi_f)$;

$B'3$: $\int_0^\infty |(\psi_f''(x) + \psi_f'(x)\psi_f(x))(H(x) - F(x))|\, dx < \infty$.

Under these conditions the asymptotic variances of $\sqrt{n}\ M^*(F)$ and $\sqrt{n}\ R(F)$ are given by:

$$\sigma_H^2(M^*(F)) = \tau_H^2\ \frac{\int\ \psi^2(x/\tau_H)\ dH(x)}{\left[\int\ \psi'(x/\tau_H)\ dH(x)\right]^2}$$

where τ_H satisfies $\int\ \psi^2(x/\tau_H)\ dH(x) = \int\ \psi^2(x)\ dF(x)$, and

$$\sigma_H^2(R(F)) = \frac{\int_0^1\ J^2(t)\ dt}{\left[\int\ J'(H(x))h^2(x)\ dx\right]^2}\ .$$

The relationship between these asymptotic variances for varying underlying distributions H is given by the following theorem.

Theorem 2: Let F satisfy conditions $A'1$-$A'4$. Then

$$\sigma_H^2(M^*(F)) \geq \sigma_H^2(R(F)) \qquad (2)$$

for all distributions H satisfying $B'1$-$B'3$ and equality in (2) holds if and only if $H(x) = F(ax)$ for some $a > 0$.

Before we prove Theorem 2 we have to state and prove several Lemmas.

Lemma 1. If θ_1, θ_2 are two nonnegative real numbers, satisfying $\theta_1 + \theta_2 \leq 1$, then

$$(\theta_1 x + \theta_2 y)^2 \leq \theta_1 x^2 + \theta_2 y^2 \qquad \text{for all } x, y \in \mathbb{R}. \tag{3}$$

If $\theta_1, \theta_2 > 0$, then equality in (3) implies $x = y$.

Proof:

$$(\theta_1 x + \theta_2 y)^2 = \theta_1 x^2 + \theta_2 y^2 - \theta_1\theta_2(x-y)^2 - (1-\theta_1-\theta_2)(\theta_1 x^2 + \theta_2 y^2)\ .$$

Lemma 2. Let φ be a function mapping the open interval $(\frac{1}{2}, 1)$ into the set of positive real numbers. Let $1/\varphi(t)$ be concave on $(\frac{1}{2}, 1)$. Then

$$\varphi(\lambda U_1 + (1-\lambda)U_2)\ (\lambda h_1 + (1-\lambda)h_2)^2 \leq \lambda\varphi(U_1)h_1^2 + (1-\lambda)\varphi(U_2)h_2^2$$

for all $0 \leq \lambda \leq 1$, $U_1, U_2 \in (\frac{1}{2}, 1)$, $h_1, h_2 \in \mathbb{R}$.

Proof: Set $V_\lambda = [\varphi(\lambda U_1 + (1-\lambda)U_2)]^{-1}$ and let $0 \leq \lambda \leq 1$. Then concavity of $1/\varphi$ implies $V_\lambda \geq \lambda V_1 + (1-\lambda)V_0$. Setting $\theta_1 = \lambda V_1/V_\lambda$ and $\theta_2 = (1-\lambda)V_0/V_\lambda$, we obtain $\theta_1 + \theta_2 \leq 1$. Thus Lemma 1 implies

$$\varphi(\lambda U_1 + (1-\lambda)U_2)(\lambda h_1 + (1-\lambda)h_2)^2 = V_\lambda \left(\frac{\lambda V_1}{V_\lambda}\ \frac{h_1}{V_1} + \frac{(1-\lambda)V_0}{V_\lambda}\ \frac{h_2}{V_0}\right)^2$$

$$\leq V_\lambda \left(\frac{\lambda V_1}{V_\lambda}\ \frac{h_1^2}{V_1^2} + \frac{(1-\lambda)V_0}{V_\lambda}\ \frac{h_2^2}{V_0^2}\right) = \lambda\varphi(U_1)h_1^2 + (1-\lambda)\varphi(U_2)h_2^2$$

Q.E.D.

Remark 1: The limits $\varphi(1) = \lim_{U\uparrow 1}\varphi(U)$ and $\varphi(\frac{1}{2}) = \lim_{U\downarrow\frac{1}{2}}\varphi(U)$ exist under the assumptions of Lemma 2, with $0 < \varphi(\frac{1}{2}),\ \varphi(1) \leq \infty$. Thus by taking limits the convexity inequality of Lemma 2 can be extended to hold for $U_1, U_2 \in [\frac{1}{2}, 1]$.

Remark 2: For $0 < \lambda < 1$ equality in the convexity inequality of Lemma 2 implies $h_1 = h_2$.

Definition 2: Let φ satisfy the assumptions of Lemma 2. Define $\mathcal{K}_\varphi$ to be the class of all distributions H which satisfy the following conditions:

i) H has a density $h(x)$ and $h(x) = h(-x)$.

ii) $A(H) = \int_0^\infty \varphi(H(x))h^2(x)\, dx < \infty$.

Lemma 3: Let $H_1, H_2 \in \mathcal{K}_\varphi$ and set $H_\lambda(x) = \lambda H_1(x) + (1-\lambda)H_2(x)$ for $0 \le \lambda \le 1$. Then

$$A(H_\lambda) \le \lambda A(H_1) + (1-\lambda)A(H_2)\ . \tag{4}$$

If $0 < \lambda < 1$, $H_1 \ne H_2$ and $\frac{1}{2} < H_2(x) < 1$ for $0 < x < \infty$, then inequality (4) is a strict inequality.

Proof: From Lemma 2 and Remark 1, we obtain for all $x \ge 0$:

$$\varphi(H_\lambda(x))h_\lambda^2(x) \le \lambda\varphi(H_1(x))h_1^2(x) + (1-\lambda)\varphi(H_2(x))h_2^2(x)\ . \tag{5}$$

Taking integrals on both sides of (5), one obtains the asserted inequality. Now let $0 < \lambda < 1$ and let x be such that $0 < x < \infty$ and $\frac{1}{2} < H_1(x), H_2(x) < 1$. Then equality in (5) implies $h_1(x) = h_2(x)$ by Lemma 2 and Remark 2. Let

$$D_1 = \{x : \tfrac{1}{2} < H_1(x) < 1\} \quad \text{and let} \quad D_2 = \{x : h_1(x) \ne h_2(x),\ x > 0\}\ .$$

If $H_2(x)$ is such that $\frac{1}{2} < H_2(x) < 1$ for $0 < x < \infty$, we see that

$$A(H_\lambda) < \lambda A(H_1) + (1-\lambda)A(H_2) \quad \text{provided that} \quad \mu(D_1 \cap D_2) \ne 0\ ,$$

where μ is Lebesgue measure on $\mathbb{R}$. But $\mu(D_1 \cap D_2) = 0$ implies $h_1(x) = h_2(x)$ a.e. $[\mu]$ for $x \ge 0$. Thus

$$A(H_\lambda) < \lambda A(H_1) + (1-\lambda)A(H_2) \quad \text{for} \quad 0 < \lambda < 1$$

$$\text{if} \quad H_1 \ne H_2 \text{ and if } H_2 \text{ satisfies } \tfrac{1}{2} < H_2(x) < 1 \text{ for } 0 < x < \infty.$$

Q.E.D.

Lemma 4: Let $\mathcal{E}$ be a convex class of distributions on $\mathbb{R}$ and let the functional $B(H)$ be convex in $H \in \mathcal{E}$, i.e.,

$$B(\lambda H_1 + (1-\lambda)H_2) \le \lambda B(H_1) + (1-\lambda)B(H_2)$$

for all $H_1, H_2 \in \mathcal{E}$ and $0 \le \lambda \le 1$. Let $H_0 \in \mathcal{E}$ be fixed and set $B_H(\lambda) = B(\lambda H + (1-\lambda)H_0)$ for $H \in \mathcal{E}$. If

$$\lim_{\lambda \downarrow 0} \frac{B_H(\lambda) - B_H(0)}{\lambda} \ge 0 \quad \text{for all } H \in \mathcal{E}$$

then $B(H) \ge B(H_0)$ for all $H \in \mathcal{E}$.

Proof: By convexity of B we have

$$\frac{B_H(\lambda) - B_H(0)}{\lambda} \le B(H) - B(H_0)$$

and the assertion follows.

Remark 3: If $B(H)$ is strictly convex at H_0, i.e.,

$$B(\lambda H + (1-\lambda)H_0) < \lambda B(H) + (1-\lambda)B(H_0)$$

for all $0 < \lambda < 1$ and all $H \in \mathcal{E}$ with $H \ne H_0$, then $B(H_0) < B(H)$ for all $H \in \mathcal{E}$ with $H \ne H_0$.

Lemma 5: Let $\mathcal{A}$ be a class of objects a with the property: if $a_1, a_2 \in \mathcal{A}$ then $\lambda a_1 + (1-\lambda)a_2$ is defined and in $\mathcal{A}$ for $0 \le \lambda \le 1$. Let $K : \mathcal{A} \longrightarrow \mathbb{R}$ be convex, i.e.,

$$K(\lambda a_1 + (1-\lambda)a_2) \le \lambda K(a_1) + (1-\lambda)K(a_2) \quad \text{for } 0 \le \lambda \le 1 \quad \text{and} \quad a_1, a_2 \in \mathcal{A}\ .$$

Fix $a_1, a_2 \in \mathcal{A}$ and set $K(\lambda) = K(\lambda a_1 + (1-\lambda)a_2)$, then $K(\lambda)$ is convex on $[0, 1]$.

Proof: Let $\nu, \lambda_1, \lambda_2 \in [0, 1]$, then

$$\begin{aligned} K(\nu\lambda_1 + (1-\nu)\lambda_2) &= K\left(\nu[\lambda_1 a_1 + (1-\lambda_1)a_2] + (1-\nu)[\lambda_2 a_1 + (1-\lambda_2)a_2]\right) \\ &\le \nu K(\lambda_1) + (1-\nu)K(\lambda_2) \end{aligned}$$

Q.E.D.

Definition 3: Let $\varphi_1 : (0, \infty) \longrightarrow \mathbb{R}$ and define

$$\mathcal{K}(\varphi, \varphi_1) = \left\{ H : \ H \in \mathcal{K}_\varphi \ \text{ and } \ \int_0^\infty |\varphi_1(x)| \, h(x) \, dx < \infty \right\}\ .$$

Lemma 6: Let φ satisfy the conditions of Lemma 2 and let $H_0 \in \mathcal{K}(\varphi, \varphi_1)$ be a strictly increasing distribution function. Let $H \in \mathcal{K}(\varphi, \varphi_1)$ be such that

$$0 \leq D_0(H) = \int_0^\infty \Big[\varphi'(H_0(x))(H(x) - H_0(x))h_0^2(x)$$

$$+ 2\varphi(H_0(x))[h(x) - h_0(x)]h_0(x) - \varphi_1(x)[h(x) - h_0(x)]\Big] \; dx < \infty \; .$$

Then

$$\lim_{\lambda \downarrow 0} \frac{B(\lambda H + (1-\lambda)H_0) - B(H_0)}{\lambda} \; \exists = D_0(H) \geq 0 \; ,$$

where

$$B(H) = \int_0^\infty \varphi(H(x))h^2(x) - \varphi_1(x)h(x) \; dx \; .$$

Proof: Let $H_\lambda(x) = \lambda H(x) + (1-\lambda)H_0(x)$ and let h_λ be the density of H_λ. By Lemma 2 we have for all $x > 0$:

$$a_x(\lambda) = \varphi(H_\lambda(x))h_\lambda^2(x) - \varphi_1(x)h_\lambda(x) \leq \lambda a_x(1) + (1-\lambda)a_x(0) \; ,$$

thus

$$\frac{a_x(\lambda) - a_x(0)}{\lambda} \leq a_x(1) - a_x(0) \; .$$

The functions $a_x(1)$ and $a_x(0)$ are integrable over the interval $(0, \infty)$ by assumption, and Lemma 5 establishes the convexity of $a_x(\lambda)$ in λ. Therefore $(a_x(\lambda) - a_x(0))/\lambda$ decreases as λ decreases to zero.

$$\frac{a_x(\lambda) - a_x(0)}{\lambda} \searrow \varphi'(H_0(x))(H(x) - H_0(x))h_0^2(x)$$

$$+2\varphi(H_0(x))(h(x) - h_0(x))h_0(x) - \varphi_1(x)(h(x) - h_0(x))$$

a.e. $[\mu]$ for $x > 0$. The assertion follows from the monotone convergence theorem. Q.E.D.

Proof of Theorem 2. Letting $H^*(x) = H(\tau_H x)$, we obtain for the ratio of the two variances:

$$\begin{aligned} \mathfrak{r}(H) &= \frac{\sigma_H^2(M^*(F))}{\sigma_H^2(R(F))} \\ &= \frac{\int \psi^2(x) \; dH^*(x)}{\int_0^1 J^2(t) \; dt} \left[\frac{\int J'(H^*(x)){h^*}^2(x) \; dx}{\int \psi'(x)h^*(x) \; dx}\right]^2 , \end{aligned}$$

with

$$H^* \quad \text{satisfying} \quad \int \psi^2(x)\, dH^*(x) = \int \psi^2(x)\, dF(x)\ .$$

The efficiency $\mathfrak{r}(H)$ is independent of the scale of H. Since $\int_0^1 J^2(t)\, dt = \int \psi^2(x)\, dF(x)$, this efficiency $\mathfrak{r}(H)$ reduces to

$$\mathfrak{r}(H) = \left(\frac{\int J'(H^*(x)) h^{*2}(x)\, dx}{\int \psi'(x) h^*(x)\, dx} \right)^2 .$$

We will show that $\mathfrak{r}(H) \geq 1$ and that equality holds if and only if $H(x) = F(ax)$ for some $a > 0$. It suffices to show that

$$\int J'(H^*(x)) h^{*2}(x)\, dx \geq \int \psi'(x) h^*(x)\, dx\ , \tag{6}$$

subject to the condition

$$\int \psi^2(x) h^*(x)\, dx = \int \psi^2(x) f(x)\, dx$$

and that (6) becomes an equality if and only if $F = H^*$. We observe that

$$J''(F(x)) f^2(x) = \psi''(x) + \psi'(x)\psi(x) \quad \text{and} \quad J'(F(x)) f(x) = \psi'(x) \ \text{ for } x > 0.$$

Condition $B'2$ implies $\int_0^\infty \psi'\, dH(x) < \infty$ and $\int_0^\infty \psi^2(x)\, dH(x) < \infty$. With this and condition $B'3$ we obtain by integration by parts:

$$\int_0^\infty J''(F(x)) f^2(x)[H(x) - F(x)] + f(x) J'(F(x))[h(x) - f(x)]\, dx$$

$$\begin{aligned}
&= \int_0^\infty [\psi''(x) + \psi'(x)\psi(x)][H(x) - F(x)] + \psi'(x)[h(x) - f(x)]\, dx \\
&= \left[\left(\tfrac{1}{2}\psi^2(x) + \psi'(x) \right) (H(x) - F(x)) \right]_0^\infty - \int_0^\infty \tfrac{1}{2}\psi^2(x)[h(x) - f(x)]\, dx \\
&= -\tfrac{1}{2} \int_0^\infty \psi^2(x)(h(x) - f(x))\, dx\ .
\end{aligned}$$

We will use Lemma 6 with the following identifications:

$$\varphi(t) = J'(t), \quad H_0(x) = F(x) \quad \text{and} \quad \varphi_1(x) = f(x) J'(F(x)) - \tfrac{1}{2}\varphi^2(x)\ .$$

Then $\varphi(t)$ satisfies the assumptions of Lemma 2 and has a continuous derivative on the interval $(\frac{1}{2}, 1)$. With the above, we obtain

$$D_0(H) = \int_0^\infty \Big(J''(F(x))f^2(x)[H(x) - F(x)] + 2f(x)J'(F(x))[h(x) - f(x)]$$

$$- [f(x)J'(F(x)) - \tfrac{1}{2}\psi^2(x)][h(x) - f(x)]\Big)\ dx = 0$$

Setting

$$B(H) = \int_0^\infty J'(H(x))h^2(x) - (\psi'(x) - \tfrac{1}{2}\psi^2(x))h(x)\ dx\ ,$$

it follows from Lemma 6 that $B(\lambda) = B(\lambda H + (1-\lambda)F)$ has derivative zero at $\lambda = 0$. Since $B(H)$ is strictly convex at F by Lemma 3, it follows from Lemma 4 and Remark 3 that

(7) $$B(F) < B(H) \quad \text{for} \quad H \neq F\ .$$

For $H = H^*$, where H^* satisfies

$$\int \psi^2(x)\ dH^*(x) = \int \psi^2(x)\ dF(x)\ ,$$

we see that (7) implies (6). Q.E.D.

2.3 Discussion of Results.

Whereas most conditions which are imposed on F are regularity conditions, there is one, $A'4$, which is not of that type. Condition $A'4$ is used quite strongly in the proof of Theorem 2, since through it one obtains certain convexity properties of the functional that is to be minimized. This enables us to bypass the variational approach, which seems to suggest that Theorem 2 may be true without condition $A'4$. On the other hand there may be some doubts as to the direct applicability of the calculus of variations to this problem, since the distributions H have certain conditions attached to them. Furthermore, the calculus of variations usually yields criteria only for local extrema and it seems to be difficult to see whether a local extremum is also a global one.

Now we will mention a few distributions F which satisfy condition $A'4$.

1. The logistic distribution $F(x) = (1 + \exp(-x))^{-1}$. Here $\psi(F^{-1}(u)) = J(u) = 2u - 1$, thus $[J'(u)]^{-1}$ is concave on the interval $(1/2, 1)$.

2. The normal distribution $F(x) = \Phi(x)$. Here, $\psi(F^{-1}(u)) = \Phi^{-1}(u) = J(u)$ with $J'(u) = [f(F^{-1}(u))]^{-1}$ and $[J'(u)]^{-1} = f(F^{-1}(u))$ is concave, as is seen by differentiation.

3. Let F be a distribution with density

$$f(x) = C \cdot \exp(-|x|^\alpha),\ 1 < \alpha \leq 2;$$

for $x > 0$ we have:

$$\psi(x) = \alpha x^{\alpha-1}, \qquad \psi'(x) = \alpha(\alpha-1)x^{\alpha-2}$$

and

$$\psi''(x) = \alpha(\alpha-1)(\alpha-2)x^{\alpha-3}.$$

Thus

$$(\psi'(x)\psi(x) + \psi''(x))(\psi'(x))^{-2} = (\alpha-1)^{-1}x + (\alpha(\alpha-1))^{-1}(\alpha-2)x^{1-\alpha}$$

is increasing in $x > 0$. Since

$$[J'(u)]^{-1} = f(F^{-1}(u))[\psi'(F^{-1}(u))]^{-1}$$

is concave if

$$-\frac{d}{du}[J'(u)]^{-1} = \frac{\psi'(F^{-1}(u))\psi(F^{-1}(u)) + \psi''(F^{-1}(u))}{(\psi'(F^{-1}(u)))^2}$$

is increasing, it follows that $A'4$ is satisfied.

In the case where F is the logistic distribution, we give a simpler proof for Theorem 2 in the Appendix B. In the case where F is the normal distribution, the estimator $M^*(F)$ is the ordinary sample mean; the result of Theorem 2 in this case was already proved by Chernoff and Savage (1958) and later by Gastwirth and Wolff (1968).

The above comparison is subject to the following criticism. There is a certain arbitrariness in the construction of $M^*(F)$ as far as the estimation of scale is concerned. In order to obtain scale invariant location estimators, many other ways of estimating scale could have been employed. Thus one could construct many scale invariant location estimators which would be asymptotically optimal for F, but if the underlying distribution of the sample is $H \neq F$, these estimators behave quite differently from each other, and the statement of Theorem 2 may no longer be true if $M^*(F)$ is replaced by any other such scale invariant estimator. We chose $M^*(F)$ because Huber (1964) showed for a particular ψ-function ψ_f that the estimator of scale involved here is in some sense minimax and hence robust over a certain class of distributions. From Definition 1 it can be seen that the estimator of scale $S_n^*(\psi)$ continues to be robust if the involved ψ-function is bounded; i.e., outlying observations do not influence the estimator $S_n^*(\psi)$ too much. This phenomenon (that the behavior of the M^*-estimator depends very strongly on the employed scale estimator) appears neither with L- and R-estimators nor in the case of $M^*(\Phi)$. $M^*(\Phi)$ will always be the sample mean, no matter which way the scale is estimated.

A counterexample to Theorem 2 is presented in Appendix C. In this example some of the regularity conditions for F are violated and so is $A'4$, but the violation of $A'4$ does not seem to be an essential feature of the example.

After comparing $R(F)$ with $L(F)$ and $M^*(F)$ and showing that as far as asymptotic variances are concerned, the first estimator is better than the latter two (at least under certain conditions), let us next consider the comparison of $L(F)$ with $M^*(F)$. In Appendix D it is shown that for $F(x) = (1+\exp(-x))^{-1}$ neither of the estimators $L(F)$ and $M^*(F)$ is better than the other; i.e., there exist distributions H_1 and H_2 such that $\sigma^2_{H_1}(L(F)) < \sigma^2_{H_1}(M^*(F))$ and $\sigma^2_{H_2}(L(F)) > \sigma^2_{H_2}(M^*(F))$. Thus it does not seem feasible to look for results like Theorem 1 and Theorem 2 in the comparison of $L(F)$ and $M^*(F)$.

3. Discussion of a Result by Mikulski.

Mikulski (1963) treated a problem similar to the estimation problem considered here in terms of testing. He investigated the following problem, first mentioned by Chernoff and Savage (1958). Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent samples from distributions $H((x-\mu)/\sigma)$ and $H((x-\mu-\Delta)/\sigma)$ respectively. If $H(x) = F(x)$, where F is a known distribution function which is sufficiently regular, "asymptotically best" linear rank tests, say φ_F, can be constructed for the hypothesis $\mathcal{H} : \Delta \leq 0$ against the alternative $\mathcal{K} : \Delta > 0$; see Chernoff and Savage (1958). We will specify below what is meant by "asymptotically best". If $F = \Phi$, where Φ is the standard normal distribution function, asymptotically best tests are the Fisher-Yates test φ_Φ and the van der Waerden x-test. When $F = \Phi$, we can also use the two sample t-test for this testing problem, the t-test being the uniformly most powerful invariant test for finite sample sizes, thus also "asymptotically best". Denote this test by t_Φ. Let $e(\varphi_\Phi, t_\Phi, H^*)$ denote the asymptotic relative efficiency (Pitman) of φ_Φ relative to t_Φ

when the underlying distributions of the samples are $H^*(x) = H((x-\mu)/\sigma)$ and $H^*(x-\Delta_N)$ respectively, with Δ_N converging to zero at a certain rate. Chernoff and Savage show that $e(\varphi_\Phi, t_\Phi, H^*) \geq 1$ for all H^*, subject to certain regularity conditions, and that equality holds if and only if $H^*(x) = \Phi((x-m)/a)$ for some m and $a > 0$.

The question they pose is: Can a similar statement be made about $e(\varphi_F, t_F, H^*)$, where φ_F is an "asymptotically best" rank test for F, and t_F is a parametric test that is "asymptotically best" for F? Mikulski shows that for any $F \neq \Phi$ satisfying certain regularity conditions, this is not possible; i.e., $e(\varphi_F, t_F, H^*) < 1$ for some H^*. There is a certain vagueness in this problem, since it is not at all clear which parametric test t_F should be used. It seems reasonable, though, to use a test t_F which shares certain properties with t_Φ, since one is interested in a comparable situation to the one for which Chernoff and Savage obtained their result. Such properties of the t-test t_Φ are the invariance of this test under a common shift and a common change in scale in both samples. One more reason to require this property is that it is a property of the rank tests, and it thus seems only fair to require it from any competing parametric test. Mikulski did not impose this restriction and in fact his best parametric test is neither location nor scale invariant. This fact plays an important role in the variational argument of his proof. Hájek (1962) recognized the role of invariance in this problem and conjectured that the answer to the proposed problem may be positive; i.e., $e(\varphi_F, t_F, H^*) \geq 1$, if t_F has these invariance properties. Under certain restrictions we shall in the following prove Hájek's conjecture.

Let us now consider independent samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from distributions $F((x-\mu)/\sigma)$ and $F((x-\nu)/\sigma)$ respectively, where F is known to us, and construct an "asymptotically best" test which is location and scale invariant. Previously we had constructed $M^\star$- and $S^\star$-estimators for the location and scale parameters of such samples. Here we will denote the corresponding estimators for μ and σ from the X-sample by $\widehat{\mu}$ and $\widehat{\sigma}_1$ and for ν and σ from the Y-sample by $\widehat{\nu}$ and $\widehat{\sigma}_2$. These estimators have the following invariance properties:

i) $\widehat{\mu}(b\boldsymbol{X}_m + a) = b\,\widehat{\mu}(\boldsymbol{X}_m) + a$ and similarly for $\widehat{\nu}(\boldsymbol{Y}_n)$;

ii) $\widehat{\sigma}_1(b\boldsymbol{X}_m + a) = |b|\,\widehat{\sigma}_1(\boldsymbol{X}_m)$ and similarly for $\widehat{\sigma}_2(\boldsymbol{X}_m)$.

To be more precise, one should write $\widehat{\nu}_F$ instead of $\widehat{\nu}$ and similarly for the other estimators, since they were constructed with reference to F, but we will drop this index F whenever no confusion will arise thereby.

In analogy to the t-test t_Φ we now propose the following test statistic

$$t_F = \frac{(\widehat{\nu} - \widehat{\mu})\sqrt{mn/N}}{\sqrt{((m-1)/(N-2))\widehat{\sigma}_1^2 + ((n-1)/(N-2))\widehat{\sigma}_2^2}}, \quad N = m+n$$

and our t_F-test will reject when t_F is too large.

In what sense is this test "asymptotically best"? Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent samples from $F((x-\mu)/\sigma)$ and $F((x-\mu-\Delta)/\sigma)$ respectively and consider the hypotheses $\mathcal{H}_F : \Delta \leq 0$ and $\mathcal{K}_F : \Delta > 0$. Let $\beta_N(\Delta)$ denote the power function of any test for the problem of testing $\mathcal{H}_F$ against $\mathcal{K}_F$. If this test is consistent, then $\beta_N(\Delta) \to 1$ as $N \to \infty$ for any $\Delta > 0$, where we assume that as $N \to \infty$, the ratio $\lambda_N = m/N \to \lambda$ with $0 < \lambda < 1$. For comparing two sequences of tests, we consider alternatives of the form

$$\Delta_N(\delta) = \delta\sqrt{N/(mn)}, \quad \delta > 0.$$

We proceed with the following definitions:

Definition 4. A two sample test φ for testing $\mathcal{H}_F$ is asymptotically level α ($0 \leq \alpha \leq 1$) if its power function $\beta_N(\Delta)$ satisfies

$$\limsup_{N\to\infty} \beta_N(\Delta) \leq \alpha \quad \text{for all} \quad \Delta \leq 0\ .$$

The class of such tests will be denoted by C_α.

Definition 5. A two sample test φ for testing $\mathcal{H}_F$ against $\mathcal{K}_F$ will be called invariant if it is invariant under a common change of location and scale in both samples. The class of such tests will be denoted by $\mathcal{I}$.

The test statistic t_F is obviously invariant.

Definition 6. A two sample test $\varphi^\star \in \mathcal{I} \cap C_\alpha$ for testing $\mathcal{H}_F$ against $\mathcal{K}_F$ is "asymptotically best at level α among invariant tests" if

$$\lim_{N\to\infty} \beta_{\varphi^\star,N}(\Delta_N(\delta)) \geq \lim_{N\to\infty} \beta_{\varphi,N}(\Delta_N(\delta))$$

for all $\varphi \in \mathcal{I} \cap C_\alpha$ and for all $\delta > 0$. Here $\beta_{\varphi,N}(\Delta)$ denotes the power of the test φ at Δ.

Such a test $\varphi^\star$ will now simply be called "asymptotically best" for F. We will show that the test t_F is asymptotically best for F, where F satisfies the regularity conditions which were imposed in the comparison of $M^\star(F)$ and $R(F)$.

First let us consider the following auxiliary problem: A) Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent samples from distributions $F^\star(x) = F((x-\mu)/\sigma)$ and $F^\star(x-\Delta)$ respectively. Let μ and σ be fixed and consider testing the simple hypothesis $\mathcal{H}_1 : \Delta = 0$ against the simple alternative $\mathcal{K}_1 : \Delta = \Delta_N(\delta)$ ($\delta > 0$ fixed). Let φ be an invariant test for this problem which is asymptotically level α and denote its power by $\beta_N(\Delta_N(\delta))$. We are interested in finding the highest possible value which the limit of $\beta_N(\Delta_N(\delta))$ (as $N \to \infty$) can achieve among all such tests φ. Since we are considering invariant tests, we can equivalently study the following modified problem: B) Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent samples from distributions $F^\star(x) = F(x + (1-\lambda_N)\Delta)$ and $F^\star(x - \lambda_N\Delta)$ respectively and test the simple hypothesis $\mathcal{H}_2 : \Delta = 0$ against the simple alternative $\mathcal{K}_2 : \Delta = \Delta_N(\delta)$ ($\delta > 0$ fixed). It is clear that any test which is invariant and asymptotically of level α for problem B) is also invariant and asymptotically of level α for problem A) and vice versa. Also the power of the test is the same in these two models. For problem B) the most powerful level α test among all tests can be constructed by means of the Neyman-Pearson lemma and it is seen that the highest asymptotic power that can be achieved by any asymptotically level α test is

$$\beta^\star(\delta) = 1 - \Phi(u_\alpha - \kappa(\delta)),$$

where u_α is such that $\Phi(u_\alpha) = 1 - \alpha$ and $\kappa^2(\delta) = (\delta/\sigma)^2 I(f)$ with

$$I(f) = \int (f'(x)/f(x))^2 \, f(x) \, dx,$$

f being the density of F. For the derivation of this result we refer to Witting and Nölle (1970), pp 66-68.

In order to show that the test t_F is asymptotically best for F, it remains to show that t_F achieves this power asymptotically. In the further study of the asymptotic behavior of t_F we shall, for the sake of simplicity, restrict ourselves to samples which arise from distributions that are symmetric around their respective medians. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be two independent samples from symmetric distributions $H(x)$ and $H(x-\Delta)$ respectively. For the following discussion it is immaterial whether δ is fixed or depends on N. We may also assume without loss of generality that H is symmetric around zero.

Under suitable regularity conditions on H (see Section 1.1) we have that:

$$\sqrt{n}\,(\widehat{\nu}_n - \Delta) \;\xrightarrow{\mathcal{L}}\; N(0, \alpha(H)/a^2(H))$$

and

$$\sqrt{m}\,\widehat{\mu}_m \xrightarrow{\mathcal{L}} N(0, \alpha(H)/a^2(H))$$

and that

$$\widehat{\sigma}_{1m} \longrightarrow \tau_H \quad \text{and} \quad \widehat{\sigma}_{2m} \longrightarrow \tau_H \quad \text{in probability as } m, n \to \infty ,$$

where τ_H is such that

$$\int \psi^2(x/\tau_H)\, dH(x) = \int \psi^2(x)\, dF(x) \qquad \text{with} \qquad \psi(x) = -f'(x)/f(x) ,$$

and

$$\alpha(H) = \int \psi^2(x/\tau_H)\, dH(x) \quad \text{and} \quad a(H) = \tau_H^{-1} \int \psi'(x/\tau_H)\, dH(x) .$$

With the convention that P_Δ denotes the probability law of the joint samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, it then follows that

$$t_F(\Delta) = \frac{\sqrt{mn/N}\ (\widehat{\nu} - \widehat{\mu} - \Delta)}{\sqrt{[(m-1)/(N-2)]\widehat{\sigma}_1^2 + [(n-1)/(N-2)]\widehat{\sigma}_2^2}}$$

$$\xrightarrow{\mathcal{L}_{P_\Delta}} N(0, \alpha(H)/[a(H)\tau_H]^2)$$

where $t_F(0) = t_F$.

Because of the location invariance we also have

$$P_\Delta(t_F(\Delta) \le t) = P_0(t_F \le t) .$$

Let $\sigma^2(H) = \alpha(H)/[a(H)\tau_H]^2$ and let the t_F test reject if $t_F \ge u_\alpha \sigma(H)$. Then for any sequence Δ_N

$$\begin{aligned} A_N(\Delta_N) &= P_{\Delta_N}(t_F \ge u_\alpha \sigma(H)) \\ &= P_0\left(t_F \ge u_\alpha\sigma(H) - \frac{\Delta_N\sqrt{mn/N}}{\sqrt{[(m-1)/(N-2)]\widehat{\sigma}_1^2 + [(n-1)/(N-2)]\widehat{\sigma}_2^2}}\right) . \end{aligned}$$

For

$$\Delta_N \equiv 0 : \quad A_N(\Delta_N) \longrightarrow 1 - \Phi(u_\alpha) = \alpha .$$

For

$$\Delta_N \leq 0: \quad \limsup_{N\to\infty} A_N\left(\Delta_N\right) \leq \alpha \; .$$

If

$$\Delta_N = \Delta_N(\delta) = \delta\sqrt{\frac{N}{mn}}, \quad \delta > 0 \; ,$$

we have

$$A_N\left(\Delta_N(\delta)\right) \longrightarrow 1 - \Phi(u_\alpha - \delta/[\sigma(H)\tau_H]) \; .$$

Thus the t_F test is asymptotically of level α and its asymptotic power at the sequence $\Delta_N(\delta)$ is $1-\Phi(u_\alpha - \delta/[\sigma(H)\tau_H])$. If $H(x) = F(x/\sigma)$ with F symmetric around zero, then $\tau_H = \sigma$,

$$\alpha(H) = \int \psi^2(x)\, dF(x) = I(f), \quad \text{and} \quad a(H) = \frac{1}{\sigma}\int \psi'(x)\, dF(x) = \frac{I(f)}{\sigma} \; ;$$

hence in this case the asymptotic power is

$$1 - \Phi\left(u_\alpha - \frac{\delta}{\sigma}\sqrt{I(f)}\right) \; ,$$

which in the light of the previous remarks establishes that t_F is asymptotically best at F.

According to Chernoff and Savage there also exists an asymptotically best linear rank test φ_F with score function $J(u) = \psi(F^{-1}(u))$. The asymptotic power of the level α φ_F-test against the sequence $\Delta_N(\delta) = \delta\sqrt{N/(mn)}$ is

$$1 - \Phi\left(u_\alpha - \frac{\delta}{\sqrt{I(f)}} \int J'(H(x))h^2(x)\, dx\right) \; .$$

Hence the asymptotic relative efficiency of φ_F relative to t_F for samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ arising from distributions $H(x)$ and $H(x - \Delta_N(\delta))$ respectively is:

$$\mathfrak{e}(\varphi_F, t_F, H) = \frac{\alpha(H)}{a^2(H)I(f)} \left(\int J'(H(x))h^2(x)\, dx\right)^2 \; .$$

This expression is the same as the one which was studied in the comparison of the estimates $R(F)$ and $M^\star(F)$ in Section 2.2. Imposing the same conditions on F and H as in that former comparison, we obtain $\mathfrak{e}(\varphi_F, t_F, H) \geq 1$ with equality if and only if $H(x) = F(ax)$ for some $a > 0$. In view of the restriction to symmetric distributions, this result is not quite as general as the one presented by Chernoff and Savage in the normal case. In conclusion we emphasize that the result does not contradict those of Mikulski. It uses a different approach to the same problem.

APPENDIX

A. Existence and Uniqueness of $M_n^\star(\psi)$ and $S_n^\star(\psi)$.

In Definition 1, Section 1.1, $M_n^\star(\psi)$ and $S_n^\star(\psi)$ were defined as the solutions of the equations (0). It remains to show that these solutions exist and are unique.

Let G be the empirical distribution function of the sample $X_1, \ldots, X_n$; the equations (0) can be written:

$$E_G\, \psi\left(\frac{x-m}{s}\right) = 0 \qquad \text{and} \qquad E_G\, \psi^2\left(\frac{x-m}{s}\right) = \beta\ ,$$

where β satisfies $0 < \beta < \sup\{\psi^2(x) :\ x \in \mathbb{R}\}$. Here E_G denotes the expectation with respect to the distribution G. Since $X_1, \ldots, X_n$ is fixed throughout this argument, we write E instead of E_G. It is assumed that ψ is continuously differentiable with $\psi'(x) > 0$ and $\psi(x) = -\psi(-x)$ for all $x \in \mathbb{R}$. Let $\varphi(m,s) : \mathbb{R} \times \mathbb{R}^+ \longrightarrow \mathbb{R} \times \mathbb{R}^+$ be the following map:

$$\varphi(m,s) = \begin{pmatrix} E\ \psi\left(\frac{x-m}{s}\right) \\ E\ \psi^2\left(\frac{x-m}{s}\right) \end{pmatrix}.$$

The Jacobian of this map is:

$$J_\varphi = -\frac{1}{s}\begin{pmatrix} E\ \psi'\left(\frac{x-m}{s}\right) & E\ \left[\frac{x-m}{s}\ \psi'\left(\frac{x-m}{s}\right)\right] \\ 2E\ \left[\psi'\left(\frac{x-m}{s}\right)\psi\left(\frac{x-m}{s}\right)\right] & 2E\ \left[\frac{x-m}{s}\ \psi'\left(\frac{x-m}{s}\right)\psi\left(\frac{x-m}{s}\right)\right] \end{pmatrix}.$$

Letting $y = (x-m)/s$, we obtain for the determinant of J_φ

$$\begin{aligned} \det\ J_\varphi &= -(2/s)\left[E\left(\psi'(y)\right)E\left(y\psi'(y)\psi(y)\right) - E\left(y\psi'(y)\right)E\left(\psi'(y)\psi(y)\right)\right] \\ &= -\,(2/s)E^2(\psi'(y))\left[\frac{E(y\psi(y)\psi'(y))}{E\psi'(y)} - \frac{E(y\psi'(y))}{E\psi'(y)}\ \frac{E(\psi(y)\psi'(y))}{E\psi'(y)}\right]. \end{aligned}$$

Since $\psi' > 0$ implies $E\psi'(y) > 0$, one can define the new probability measure $G^\star$ by

$$dG^\star = \frac{\psi'}{E_G\ \psi'(y)}\ dG$$

so that

$$\begin{aligned} \det\ J_\varphi &= -(2/s)E^2(\psi'(y))\left[E_{G^\star}(y\psi(y)) - E_{G^\star}(y)E_{G^\star}(\psi(y))\right] \\ &= -\,(2/s)E^2\psi'(y)\ \mathrm{cov}_{G^\star}(y,\psi(y)) < 0\ . \end{aligned}$$

The last inequality follows from

Lemma 7: If $EY^2 < \infty$ and $E\psi^2(Y) < \infty$ and if $\psi(x) > \psi(y)$ for $x > y$, then $\text{cov}(Y, \psi(Y)) > 0$ for any distribution of Y that does not concentrate on one point.

Proof: Let Y_1, Y_2 be two independent identically distributed random variables, then $(Y_1 - Y_2)(\psi(Y_1) - \psi(Y_2)) > 0$ for $Y_1 \neq Y_2$. Since $P(Y_1 \neq Y_2) > 0$, we have

$$E[(Y_1 - Y_2)(\psi(Y_1) - \psi(Y_2))] > 0 \ , \quad \text{hence} \quad \text{cov}(Y_1, \psi(Y_1)) > 0$$

Q.E.D.

Thus the Jacobian of φ has a determinant which is always negative and so are the two main diagonal elements of the Jacobian. Theorem 4 of a paper by Gale and Nikaidô (1965) enables us to conclude that φ is a one to one map. This establishes the uniqueness of the solutions of the above equations. In order to prove the existence of the solutions we observe that for each s there exists an $m(s)$ such that

$$\frac{1}{n}\sum_{i=1}^{n} \psi\left(\frac{X_i - m(s)}{s}\right) = 0 \ ,$$

and $m(s)$ satisfies: $\min_{1\le i\le n} X_i \le m(s) \le \max_{1\le i\le n} X_i$. As s varies from zero to infinity, it is seen that

$$\frac{1}{n}\sum_{i=1}^{n} \psi^2\left(\frac{X_i - m(s)}{s}\right)$$

ranges from $\sup\{\psi^2(x) : \ x \in \mathbb{R}\}$ to zero. Q.E.D.

B. A simpler Proof of Theorem 2 for the Case that *F* is the Logistic Distribution.

Let $F(x) = (1 + \exp(-x))^{-1}$; then $f(x) = \exp(-x)(1 + \exp(-x))^{-2}$ and

$$\psi_F(x) = \psi(x) = -\frac{f'(x)}{f(x)} = \frac{1 - \exp(-x)}{1 + \exp(-x)} \ .$$

One observes the following relations: $\psi'(x) = (1 - \psi^2(x))/2 = 2f(x)$ and $J(t) = \psi(F^{-1}(t)) = 2t - 1$.

Let $h_0(x)$ be a density satisfying

$$\int \psi^2(x)h_0(x)\,dx = \int \psi^2(x)f(x)\,dx \ \ (= 1/3) \ .$$

Since

$$3\int \psi'(x)h_0(x)\,dx = 3\int \tfrac{1}{2}(1-\psi^2(x))h_0(x)\,dx = 1\ ,$$

we obtain by Jensen's inequality

$$\begin{aligned}\int h_0^2(x)dx &= \int \frac{h_0(x)}{3\psi'(x)}\,3\psi'(x)h_0(x)\,dx \\ &\geq \frac{1}{9}\frac{1}{\int (\psi'(x))^2\,dx} = \frac{1}{9}\frac{1}{\int 4f^2(x)\,dx} = \frac{1}{6}\ .\end{aligned}$$

The ratio of the asymptotic variances of $M^\star(F)$ and $R(F)$ is

$$\frac{\sigma_H^2(M^\star(F))}{\sigma_H^2(R(F))} = \frac{\left(\int 2h^2(x)\,dx\right)^2}{\int_0^1 J^2(t)\,dt}\,\frac{\int \psi^2(x/\tau_H)\,dH(x)}{\left(\int \psi'(x/\tau_H)\,dH(x)\right)^2}\,\tau_H^2\ ,$$

where τ_H satisfies $\int \psi^2(x/\tau_H)\,dH(x) = \int \psi^2(x)\,dF(x)$. Setting $H_0(x) = H(x\tau_H)$, we have $\int \psi^2(x)\,dH_0(x) = \int \psi^2(x)\,dF(x)$ and the above inequality implies

$$\frac{\sigma_H^2(M^\star(F))}{\sigma_H^2(R(F))} = \frac{\left(\int 2h_0^2(x)\,dx\right)^2}{\int_0^1 J^2(t)\,dt}\,\frac{\int \psi^2(x)\,dH_0(x)}{\left(\int \psi'(x)\,dH_0(x)\right)^2} \geq 1\ .$$

It is also seen that equality holds if and only if $H(x) = F(ax)$ for some $a > 0$.

C. A Counter Example to Theorem 2.

Let

$$f(t) = (1-\epsilon)\frac{1}{\sqrt{2\pi}}\exp(-\rho(t))\ ,$$

where

$$\rho(t) = \begin{cases} t^2/2 & \text{for } |t| \leq k \\ k|t| - k^2/2 & \text{for } |t| > k \end{cases}$$

and where ϵ and k satisfy the relation

$$\frac{1}{1-\epsilon} = 2\Phi(k) - 1 + 2\varphi(k)/k\ .$$

Here Φ is the standard normal distribution function and φ its density. Then

$$\psi_f(x) = \psi(x) = \begin{cases} x & \text{for } |x| \leq k \\ k\,\text{sign}(x) & \text{for } |x| > k \end{cases}$$

The corresponding estimators $M^\star(F)$ are those of Huber's (1964) proposal two. Conditions for their asymptotic normality are given by Huber (1967).

Let $R(F)$ be the corresponding R-estimator. It was seen in the proof of Theorem 2 that the statement:

$$\sigma_H^2(M^\star(F)) \geq \sigma_H^2(R(F)) \quad \text{for all } H \text{ in some class} \tag{8}$$

is equivalent to:

$$\int \frac{d}{dx}\left[\psi\left(F^{-1}(H(x))\right)\right]\, dH(x) \geq \int \psi'(x)\, dH(x) \quad \text{for all } H \text{ in that class} \tag{9}$$

which satisfy the constraint

$$\int \psi^2(x)\, dH(x) = \int \psi^2(x)\, dF(x)\ . \tag{10}$$

It will be shown that there exists an H_0 satisfying (10) for which the inequality (9) is not satisfied.

Since we consider only distributions H which have a density $h(x)$ and which are symmetric around zero one may write (9) and (10) as follows:

$$\int_0^\alpha [F^{-1}(H(x))]'\, dH(x) \geq H(k) - 1/2\ , \quad \text{where } \alpha = H^{-1}(F(k)) \tag{11}$$

and

$$\int_0^k x^2 h(x)\, dx + k^2(1 - H(k)) = \int_0^k x^2 f(x) dx + k^2(1 - F(k)) \tag{12}$$

or equivalently after integration by parts

$$\int_0^k x(H(x) - F(x))\, dx = 0\ . \tag{13}$$

Let K_α be the class of distributions H that satisfy the following conditions:

i) H has a density $h(x)$ with $h(x) = h(-x)$;

ii) $H \in CR(J) \cap CM^\star(\psi)$ with $J(t) = \psi(F^{-1}(t))$;

iii) $\int_0^k x(H(x)-F(x))\,dx = 0$ and $H(\alpha) = F(k)$.

K_α is a convex class of distributions. Assume $\alpha \geq k$ and let

$$C(H) = \int_0^\alpha [F^{-1}(H(x))]'\,dH(x) \qquad \text{for } H \in K_\alpha$$

and set $C(\lambda) = C(\lambda H_1 + (1-\lambda)H_2)$, where $H_1 \in K_\alpha$ and H_2 be the distribution function corresponding to the density:

$$h_2(x) = \begin{cases} f(x) & \text{for } |x| \leq k \\ 0 & \text{for } k < |x| \leq \alpha \end{cases}$$

and complete the definition of h_2 in such a way that h_2 is a symmetric bounded density. Thus we also have $H_2 \in K_\alpha$. Let

$$H_\lambda(x) = \lambda H_1(x) + (1-\lambda)H_2(x) \qquad \text{and} \qquad h_\lambda(x) = \lambda h_1(x) + (1-\lambda)h_2(x).$$

We will study the derivative of $C(\lambda)$ with respect to λ at $\lambda = 0$.

$$\begin{aligned} C(\lambda) - C(0) &= \int_0^\alpha J'(H_\lambda(x))h_\lambda^2(x)\,dx - \int_0^\alpha J'(H_2(x))h_2^2(x)\,dx \\ &= \int_0^k \left[J'(H_\lambda(x))h_\lambda^2(x) - J'(H_2(x))h_2^2(x)\right]\,dx \\ &\qquad\qquad + \int_k^\alpha J'(H_\lambda(x))\lambda^2 h_1^2(x)\,dx\ . \end{aligned}$$

The first summand on the right shall be denoted by $B(\lambda)$ and the second by $A(\lambda)$. Then

$$\frac{A(\lambda)}{\lambda} \longrightarrow 0 \quad \text{as} \quad \lambda \to 0$$

and

$$\begin{aligned} \frac{B(\lambda)}{\lambda} \overset{\lambda\to 0}{\longrightarrow} & \int_0^k \Big[J''(H_2(x))(H_1(x)-H_2(x))h_2^2(x) \\ & \qquad + J'(H_2(x))\cdot 2\cdot h_2(x)(h_1(x)-h_2(x))\Big]\,dx \\ &= \int_0^k x(H_1(x)-F(x)) + 2(h_1(x)-f(x))\,dx \\ &= 2(H_1(k)-F(k)) \qquad \text{since} \qquad H_1 \in K_\alpha\ . \end{aligned}$$

Thus

$$\left.\frac{C(\lambda)}{d\lambda}\right|_{\lambda=0} = 2(H_1(k) - F(k)) \ .$$

Let $C^\star(H) = C(H) - H(k) + 1/2$ and $C^\star(\lambda) = C^\star(H_\lambda)$. We observe that $C^\star(0) = 0$ and

$$C^{\star\prime}(0) = \left.\frac{dC^\star(\lambda)}{d\lambda}\right|_{\lambda=0} = H_1(k) - F(k) \ .$$

If there exists an $H_1 \in K_\alpha$ (for $\alpha > k$) which satisfies $H_1(k) < F(k)$, then $C^{\star\prime}(0) < 0$. This implies that $C^\star(\lambda) < 0$ for some $\lambda > 0$, which in turn implies that there exists an $H_0 \in K_\alpha$ $(\alpha > k)$ with $\sigma^2_{H_0}(M^\star(F)) > \sigma^2_{H_0}(R(F))$. It remains to show that there exists an $H_1 \in K_\alpha$ $(\alpha > k)$ with $H_1(k) < F(k)$. Since we have to satisfy $\int_0^k x(H_1(x) - F(x))\ dx = 0$, we can restrict attention to the interval $[0, k]$. It is easily seen that $\int_0^k 2xH_1(x)\ dx$ can take on any value in the interval $(k^2/2, H_1(k)k^2)$, as H_1 varies on the interval $[0, k]$ subject to the conditions $H_1(0) = 1/2$ and $H_1(k) < F(k)$. Since $k^2/2 < \int_0^k 2xF(x)\ dx < F(k)k^2$, one can clearly find a distribution H_1 with bounded density such that $H_1(k) < F(k)$ and $\int_0^k x(F(x) - H_1(x))\ dx = 0$. Q.E.D.

D. Examples for the Comparison of $M^\star(F)$ and $L(F)$.

Consider the case of the logistic distribution $F(x) = (1+\exp(-x))^{-1}$ and observe the following relations:

$$\psi(x) = -f'(x)/f(x) = \tanh(x/2) = \frac{1-\exp(-x)}{1+\exp(-x)}, \qquad \psi(F^{-1}(t)) = 2t - 1,$$

$$\psi'(F^{-1}(t)) = 2t(1-t), \quad I(f) = \int \psi^2(x)\ dF(x) = 1/3 \ \text{ and } \ \psi'(x) = (1 - \psi^2(x))/2.$$

We have $\sigma^2_H(L(F)) = \int_0^1 U^2(t)\ dt$, where

$$U(t) = \int_{1/2}^t \left[\psi'(F^{-1}(u))/h(H^{-1}(u))\right]\ du\ \ [I(f)]^{-1} \ .$$

Let $H(x) = H_a(x)$ be the uniform distribution on the interval $[-a, +a]$, then

$$U(t) = 4a \int_{1/2}^t u(1-u)\ du = 4a(t^2/2 - t^3/3 - 1/12)$$

and thus $\sigma_H^2(L(F)) = a^2\ 17/35$. Now we choose a such that $\int\ \psi^2(x)\ dH_a(x) = \int\ \psi^2(x)\ dF(x)$; i.e., let $a = a_0$ where $\psi(a_0)/a_0 = 1/3$. Using tables of the tanh-function, one obtains $\psi(2.56)/2.56 > 1/3$, hence $a_0 > 2.56$. Since $\sigma^2_{H_{a_0}}(M^\star(F)) = 3$, we have

$$\sigma^2_{H_{a_0}}(M^\star(F)) = 3 < (2.56)^2\ 17/35 < \sigma^2_{H_{a_0}}(L(F))\ .$$

Now let H_a be the double exponential distribution; i.e., H_a has a density $h_a(x) = (a/2)\exp(-a|x|)$. Determine a_0 such that

$$\int\ \psi'(x)\ dH_{a_0}(x) = \int\ \psi^2(x)\ dF(x) = 1/3\ .$$

For $a = 1$ one has

$$\int\ \psi'(x)\ dH_a = 2\log 2 - 1 > 1/3\ ,$$

hence $a_0 > 1$. Since $h_a(H_a^{-1}(t)) = (1-t)a$ for $t \geq 1/2$, we have

$$U(t) = 3\int_{1/2}^{t} \frac{2u(1-u)}{(1-u)a}\ du = (3/a)(t^2 - 1/4)$$

for $t \geq 1/2$ and hence obtain for $a = a_0 > 1$

$$\int_0^1\ U^2(t)\ dt = \frac{2}{a_0^2}\ 9\int_{1/2}^1 (t^2 - 1/4)^2\ dt = \frac{57}{40a_0^2} < 3\ ,$$

which implies $\sigma^2_{H_{a_0}}(L(F)) < \sigma^2_{H_{a_0}}(M^\star(F))$. Thus neither of the two estimates $L(F)$ and $M^\star(F)$ is uniformly better than the other as far as asymptotic variances are concerned.

References

Chernoff, H., J.L. Gastwirth and M.V. Johns (1967), "Asymptotic distributions of linear combinations of functions of order statistics with applications to estimation," *Ann. Math. Statist.* **38**, 52-72.

Chernoff, H. and I.R. Savage (1958), "Asymptotic normality and efficiency of certain nonparametric test statistics," *Ann. Math. Statist.* **29**, 972-994.

Gale, D. and H. Nikaidô (1965), "The Jacobian matrix and global univalence of mappings," *Mathematische Annalen*, 159. Band, 2. Heft, 81-93.

Gastwirth, J.L. and S. Wolff (1968), "An elementary method for obtaining lower bounds on the asymptotic power of rank tests," *Ann. Math. Statist.* **39**, 2128-2130.

Hájek, J. (1962), "Asymptotically most powerful rank-order tests," *Ann. Math. Statist.* **33**, 1124-1147.

Hájek, J. (1971), "Limiting properties of likelihoods and inference," *Proc. Symp. Foundations of Statistical Inference*, ed. by V.P. Godambe and D.A. Sprott, Holt Rinehart and Winston of Canada, Toronto, Montreal.

Hodges, J.L. and E.L. Lehmann (1963), "Estimates of location based on rank tests," *Ann. Math. Statist.* **34**, 598-611.

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

Huber, P.J. (1967), "The behavior of maximum likelihood estimates under nonstandard conditions," *Proc. Fifth Berkeley Symp. Math. Stat. and Prob.*, Vol. 1, 221-233.

Huber, P.J. (1970), 'Studentizing robust estimates," *Internat. Symp. on Nonparametric Techniques in Statistical Inference*, ed. by M.L. Puri, Cambridge University Press, 453-463.

Jaeckel, L.A. (1971), "Robust estimates of location: symmetry and asymmetric contamination," *Ann. Math. Statist.* **42**, 1020-1034.

LeCam, L. (1953), "On some asymptotic properties of maximum likelihood estimates and related Bayes estimates," *Univ. California Publ. Statist.* **1**, 277-330.

Mikulski, P. (1963), "On the efficiency of optimal nonparametric procedures in the two sample case," *Ann. Math. Statist.* **34**, 22-32.

Puri, M.L. and P.K. Sen (1971), *Nonparametric Methods in Multivariate Analysis*, John Wiley and Sons, Inc.

Witting, H. and G. Nölle (1970), *Angewandte Mathematische Statistik*, B.G. Teubner, Stuttgart.