Revue Belge de Statistique, d'Informatique et de Recherche Opérationnelle, Vol **22**, *n°* 3
Belgisch Tijdschrift voor Statistiek, Informatica en Operationeel Onderzoek, Vol. **22**, *Nr* 3

ESTIMATION AND TESTING BY MEANS OF OPTIMALLY ROBUST STATISTICS

Peter ROUSSEEUW

Vrije Universiteit Brussel

Abstract

It is showed that Hampel's optimal robustification method not only applies to M-estimators, but also to L-estimators and R-estimators. Both location and scale models are considered. By means of influence curves of general statistics, these results are extended to the corresponding tests on one or two samples. The purpose of the present paper is to guide the user of statistical methods by selecting, in each of these classes of statistical procedures, those techniques that are optimally robust.

1. INTRODUCTION: HAMPEL'S OPTIMALITY LEMMA

Let the parameter space Θ be a convex open subset of $\mathbb{R}$. The statistical model $\{F_\theta;\theta\in\Theta\}$ consists of cumulative distribution functions (cdf) on $\mathbb{R}$ (which we identify with the corresponding probability distributions), with absolutely continuous and strictly positive densities $f_\theta(x)$ with respect to the Lebesgue measure λ. We assume that $f_\theta(x)$ is differentiable with respect to θ, and we denote the derivative by $f'_\theta(x)$. Let θ_0 belong to Θ, and put $F:=F_{\theta_0}$ where the Fisher information

$$I(F) := \int (f'_{\theta_0}/f_{\theta_0})^2 dF_{\theta_0} \tag{1.1}$$

satisfies $0 < I(F) < \infty$, and where $\int f'_{\theta_0}(x)d\lambda(x) = 0$. (The latter equality is only a regularity condition, since it holds if the order of integration and differentiation may be interchanged.) A standard example of such a parametric model is $F_\theta(x)=\Phi(x-\theta)$, where Φ is the standard normal distribution, $\Theta=\mathbb{R}$ and $\theta_0=0$.

We shall now state Hampel's optimality lemma, as it was published in ([1], page 391).

<u>Lemma (Hampel)</u> Suppose the above conditions hold, and let $b>0$ be some constant. Then there exists a real number a such that

$$\tilde{\chi}(x) := [(f'_{\theta_0}(x)/f_{\theta_0}(x))-a]_{-b}^{b} \tag{1.2}$$

(where the notation means truncation at the levels b and $-b$) satisfies $\int\tilde{\chi}dF_{\theta_0} = 0$ and $c := \int\tilde{\chi}f'_{\theta_0}d\lambda > 0$. Now $\tilde{\chi}$

minimizes

$$\int \chi^2 dF_{\theta_0} \, / [\int \chi f'_{\theta_0} d\lambda]^2 \tag{1.3}$$

among all mappings χ that satisfy

$$\int \chi dF_{\theta_0} = 0 \tag{1.4}$$

$$\int \chi f'_{\theta_0} d\lambda \neq 0 \tag{1.5}$$

$$\sup_{x \in \mathbb{R}} |\chi(x) / \int \chi f'_{\theta_0} d\lambda| \leq k := \frac{b}{c} \quad . \tag{1.6}$$

This lemma was introduced in order to find optimally robust M-estimators (see section 2.2). It has indeed been argued that M-procedures should be most appropriate for robustness theory, and much work has been concentrated there. However, it is clear that procedures based on ranks (R-procedures for short) have their own advantages in view of nonparametric theory, even when efficiency is considered ([2]). On the other hand, procedures based on linear combinations of order statistics (L-procedures) are easy to perform, and some of them turned out to be favourable in another robustness context ([3]). Therefore, the extension of known results to these cases appears to be of interest.

2. OPTIMALLY ROBUST ESTIMATORS

2.1 Definition of the influence curve

Consider again the above parametric model. Our aim is to estimate the unknown parameter θ. An "estimator" is actually a whole sequence of statistics $T_N(x_1, \ldots, x_N)$, one for each sample size N. Suppose there exists a functional $T: \mathcal{M}_1(\mathbb{R}) \rightarrow \mathbb{R}$

(where $\mathcal{M}_1(\mathbb{R})$ is the space of all signed measures with mass 1 on $\mathbb{R}$), such that $T_N(x_1,\ldots,x_N)$ converges in probability to $T(G)$ when $N\to\infty$ and the observations are independently identical distributed (iid) according to the true underlying distribution G.

We shall suppose that all estimators appearing in this section are <u>Fisher-consistent</u>, that is

$$T(F_\theta) = \theta \qquad \text{for all } \theta \text{ in } \Theta. \tag{2.1}$$

In this case, the influence curve (IC) has been defined by Hampel ([1]). Denote by δ_x the Dirac probability measure in x, for all x in $\mathbb{R}$. For all ε, we construct $F_{\varepsilon,x} = (1-\varepsilon)F + \varepsilon\delta_x$ in $\mathcal{M}_1(\mathbb{R})$. The influence curve of the Fisher-consistent functional T at F is defined by

$$IC(T;F;x) := \lim_{\varepsilon\to 0}[T(F_{\varepsilon,x})-T(F)]/\varepsilon$$

for all x in which the limit exists. (Hampel originally used a right hand limit, but later it was replaced by the present two-sided limit for mathematical convenience.)

The IC describes the influence of outliers in the sample on the value of the estimator. Therefore, Hampel ([1]) introduced the <u>gross-error-sensitivity</u>

$$\gamma^*(T,F) := \sup_{x\in\mathbb{R}} |IC(T;F;x)| \quad . \tag{2.2}$$

A <u>bounded</u> IC leads to a finite sensitivity to outliers, which is a desirable feature. Under sufficient regularity conditions (implying asymptotic normality), it holds that

$$\int IC(T;F;x)^2 dF(x) = \sigma^2 \tag{2.3}$$

where σ^2 is the asymptotic variance of the estimator. The asymptotic efficiency e at F equals $(\sigma^2 I(F))^{-1}$. Therefore, one attempts to make σ^2 as small as possible. Hampel's optimality lemma will enable us to construct estimators which minimize σ^2 under the side condition of an upper bound on the gross-error-sensitivity γ^*.

2.2 M-estimators

From here on, we shall restrict ourselves to the location and scale models. In the location model we have $F_\theta(x)=F(x-\theta)$ with $\theta_0=0$. A location M-estimator ([4]) is defined by means of the implicit equation

$$\Sigma_{i=1}^{N}\psi(x_i,T_N) = 0 \quad , \tag{2.4}$$

where $\psi(x,t)=\chi(x-t)$. Here χ is a piecewise continuous and piecewise differentiable mapping, which satisfies

$$\int\chi dF = 0 \tag{2.5}$$

(reflecting Fisher-consistency), and

$$\int\chi' dF = \int\chi f'_{\theta_0} d\lambda \neq 0 \tag{2.6}$$

(which is a regularity condition). Obviously, ψ and χ are only defined up to a factor. For the IC, we obtain ([4])

$$IC(T;F;x) = \frac{\chi(x)}{\int\chi' dF} \quad , \tag{2.7}$$

which we can also write as

$$IC(T;F;x) = \frac{\chi(x)}{\int\chi f'_{\theta_0} d\lambda} \quad . \tag{2.8}$$

Now we are in a position to apply Hampel's optimality lemma.

The mapping $\tilde{\chi}$, determined by (1.2), defines a location M-estimator which is Fisher-consistent (1.4) and for which the IC exists (1.5). Its asymptotic variance σ^2, which is given by (1.3) because of (2.3), is minimal for a given upper bound k on the gross-error-sensitivity γ^*. We say that $\tilde{\chi}$ determines an <u>optimally robust</u> M-estimator. At the standard normal ($F=\Phi$) we find $(f'_{\theta_0}(x)/f_{\theta_0}(x)) = x$, so $a=0$. We then obtain

$$\tilde{\chi}(x) = [x]_{-b}^{b} \quad , \tag{2.9}$$

which is called a <u>Huber estimator</u> ([5]). If the condition on γ^* were abandoned, then we would simply obtain the maximum likelihood estimator given by $\chi_0(x) = (f'_{\theta_0}(x)/f_{\theta_0}(x))$, which in this case equals the arithmetic mean. From a robustness point of view, the latter estimator is not acceptable because $\gamma^* = \infty$.

The scale model is given by $F_\theta(x) = F(x/\theta)$, where $\Theta=\mathbb{R}_*^+$ and $\theta_0=1$. An M-estimator of scale is also given by (2.4), but now $\psi(x,t)=\chi(x/t)$. We suppose the mapping χ is still piecewise continuous, piecewise differentiable and satisfies (2.5), but now (2.6) is replaced by

$$\int y\chi'(y)dF(y) = \int \chi f'_{\theta_0} d\lambda \neq 0 \quad . \tag{2.10}$$

(Note that the expression f'_{θ_0} has different interpretations in (2.6) and (2.10).) One obtains

$$IC(T;F;x) = \frac{\chi(x)}{\int y\chi'(y)dF(y)} \quad , \tag{2.11}$$

which equals the expression (2.8) because of (2.10). Again we can apply Hampel's optimality lemma. At the standard normal, we obtain the estimator given by

$$\tilde{\chi}(x) = [x^2 - 1 - a]_{-b}^{b} \quad , \qquad (2.12)$$

which also goes back to [5].

2.3 R-estimators

One-sample R-estimators of location are described in [4]. They are based upon a score-generating function $\phi:[0,1]\to\mathbb{R}$ (sometimes denoted by J) which is only defined up to a factor, and which satisfies $\phi(1-t) = -\phi(t)$ for all t. We assume that ϕ is piecewise continuous and piecewise differentiable. The influence curve of the corresponding estimator is given in [4]. Defining the transformation $U(x) := \int_{[0,x]}\phi'[\frac{1}{2}(F(y)+1-F(2T(F)-y))]f(2T(F)-y)d\lambda(y)$ and putting $\chi(x) = U(x)-\int U(t)dF(t)$, we see that the IC equals (2.7). Therefore, Hampel's lemma can be applied: if there exists an admissible $\tilde{\phi}$ to which a χ corresponds which is a multiple of $\tilde{\chi}$, then $\tilde{\phi}$ determines an optimally robust R-estimator. At the standard normal ($F=\Phi$) we obtain

$$\tilde{\phi}(t) = [\Phi^{-1}(t)]_{-b}^{b} \quad , \qquad (2.13)$$

which is a truncated normal scores function. This estimator is similar, but never identical to Jaeckel's minimax solution ([6]), as a simple verification shows. One-sample R-estimators for the scale model do not exist.

2.4 L-estimators

Location L-estimators are of the form

$$T_N(x_1,\dots,x_N) = \Sigma_{i=1}^N a_i x^{(i)} \quad , \qquad (2.14)$$

where $x^{(1)} \leq \dots \leq x^{(N)}$ is the ordered sample. The weights a_i are generated by the formula $a_i = \int_{[(i-1)/N,i/N]} h d\lambda \,/\, \int_{[0,1]} h d\lambda$, where $h:[0,1]\to\mathbb{R}$ is Lebesgue integrable and satisfies $\int_{[0,1]} h d\lambda \neq 0$. Clearly, h is only defined up to a nonzero factor. (Therefore, one sometimes works with the standardized version $J(u) = h(u)/\int_{[0,1]} h d\lambda$; see [7].) Under diverse sets of regularity conditions (for a survey, see [8], page 15) these estimators are asymptotically normal. By differentiation of the functional

$$T(G) = \int x h(G(x)) dG(x) / \int h(F(y)) dF(y) \quad ,$$

one calculates the influence curve ([7]). Performing the substitution

$$\chi(x) := \int_{[0,x]} h(F(y)) d\lambda(y) - \int [\int_{[0,t]} h(F(y)) d\lambda(y)] dF(t) \quad ,$$

one again obtains (2.7). Its denominator equals $\int h d\lambda \neq 0$, and we still have (1.4). This enables us to apply Hampel's optimality lemma. If an admissible mapping $\tilde{h}$ exists for which the corresponding χ is a multiple of $\tilde{\chi}$, then $\tilde{h}$ determines an optimally robust L-estimator. At the normal ($F=\Phi$) we obtain the <u>trimmed mean</u>

$$\tilde{h} = \tilde{\chi}' \circ \Phi^{-1} = 1_{[\alpha,1-\alpha]} \quad , \qquad (2.15)$$

with $\alpha=\Phi(-b)$. This simple estimator is quite popular because it appeals to the intuition: one removes a fraction

α of the smallest and the largest observations, and one calculates the mean of the remaining ones. If we let b tend to zero, then the gross-error-sensitivity $\gamma^* = k$ approaches its lower bound $(\pi/2)^{1/2} \simeq 1.253$ and we obtain the <u>median</u>.

L-estimators of scale are defined by T_N of the form (2.14), but now the scores are generated by the formula $a_i = \int_{[(i-1)/N,i/N]} h d\lambda / \int_{[0,1]} h(t)F^{-1}(t)d\lambda(t)$, so

$$T(G) = \int x h(G(x))dG(x)/\int y h(F(y))dF(y)$$

where h satisfies $\int y h(F(y))dF(y) \neq 0$. Performing the same transformation of h into χ one obtains (2.11), and our earlier result (2.12) can be translated. At the standard normal, the optimal $\tilde{\chi}$ bring about the solutions

$$\begin{aligned} \tilde{h}(t) &= \Phi^{-1}(t) && \text{for } |(\Phi^{-1}(t))^2-1-a| \leq b \\ &= 0 && \text{elsewhere} \end{aligned} \qquad (2.16)$$

for suitable a and b. If b tends to zero, then $\gamma^* = k$ tends to its lower bound $(4\Phi^{-1}(0.75)\nu(\Phi^{-1}(0.75)))^{-1} \simeq 1.17$ (where ν is the normal density), and we find the <u>interquartile range estimator</u>.

3. OPTIMALLY ROBUST TESTS

3.1 Definition of the influence curve

Making use of the same model $\{F_\theta;\theta\in\Theta\}$, we now want to test the null hypothesis $\theta = \theta_0$ against the two-sided alternative $\theta \neq \theta_0$, or against a one-sided alternative ($\theta > \theta_0$ or $\theta < \theta_0$). As in subsection 2.1, one defines a sequence of statistics $T_N(x_1,\ldots,x_N)$ which converges to a functional T. One compares the test statistic T_N with one or more critical values, which depend on the level. (In practice, one often works with $N^{1/2}T_N$ or NT_N, as is the case with rank statistics.) However, T is not necessarily Fisher-consistent any more (condition (2.1)); for example, by rank tests it is not.

In order to extend the definition of the IC to this case ([9]), one introduces the mappings $\xi_N:\Theta\to\mathbb{R}$ defined by $\xi_N(\theta):=E_\theta[T_N]$, and one assumes that they converge pointwise to the mapping ξ, defined by $\xi(\theta):=T(F_\theta)$. It is also assumed that ξ is differentiable and strictly monotone, so ξ^{-1} exists. Now define U(G) as $\xi^{-1}(T(G))$; this new functional U is clearly Fisher-consistent. The influence curve of T at F is then defined as Hampel's IC of U at F. (Note that this new definition generalizes the previous one.) In case $\xi'(\theta_0)\neq 0$, it can be calculated by

$$IC(T;F;x) = \frac{\partial}{\partial\varepsilon}[T(F_{\varepsilon,x})]_{\varepsilon=0} \;/\; \xi'(\theta_0) \quad . \qquad (3.1)$$

The definition and the interpretation of the gross-error-sensitivity (2.2) remain the same.

The performance of a test is measured by its asymptotic (Pitman) efficacy c^2 , which equals

$$c^2 = \lim_{N\to\infty} [\xi_N'(\theta_0)]^2/N\, Var_{\theta_0}[T_N] \qquad (3.2)$$

under certain regularity conditions ([10]). From Cramér-Rao it follows that $0 \leq c^2 \leq I(F)$, so one can define the asymptotic efficiency e as $c^2/I(F)$. Under certain regularity conditions, it holds that ([9])

$$\int IC(T;F;x)^2 dF(x) = (c^2)^{-1} \quad . \tag{3.3}$$

We conclude that the expected square of the IC plays the same role as in the preceding section, as does γ^*. This enables us to apply Hampel's optimality lemma in this case also.

3.2 Tests of the types M and L

In section 2 we have met Fisher-consistent estimators T_N of the types M and L. These statistics can also be used for the testing problem. Asymptotic normality of M-estimators has been showed by Huber ([5]). Under diverse sets of regularity conditions, L-estimators are also asymptotically normal ([8], page 15). Therefore, if one wants to test the null hypothesis $\theta=\theta_0$, it is natural to use the test statistics

$$\sqrt{N}\,(T_N - \theta_0) \quad , \tag{3.4}$$

where T_N is an M-estimator or an L-estimator. In the location problem one works with $\sqrt{N}\,T_N$, and in the scale case one uses $\sqrt{N}\,(T_N-1)$. At the null hypothesis, the limiting distributions of these statistics are normal with mean zero. Therefore, one only has to compile tables for the critical values up to a certain N, whereafter the asymptotics take over. The optimally robust statistics T_N of the types M and L were already found in sections 2.2 and 2.4, so we only have to insert them in (3.4).

3.3 Location rank tests

One-sample rank tests only exist for the location problem ([11]). They test the hypothesis of symmetry, so we suppose that F is symmetric about zero. Let R_i^+ be the rank of $|x_i|$ in the sample $x_1,\ldots,x_N$. Rank statistics are of the form

$$T_N = \frac{1}{N} \Sigma_{i=1}^N \; a_N^+(R_i^+) \; \mathrm{sign}(x_i)$$

(in practice one often uses NT_N). The scores $a_N^+(i)$ are generated by a function $\phi^+:[0,1]\to\mathbb{R}$ which is square integrable and nondecreasing. The statistics T_N converge to

$$T(G) = \int\phi^+(G(|x|)-G(-|x|))\mathrm{sign}(x)dG(x)$$

which is not Fisher-consistent, and it holds that

$$IC(T;F;x) = \phi^+(2F(|x|)-1)\mathrm{sign}(x)/\int_0^1\phi^+(u)\phi^+(u,f)du$$

(see [9]). Here $\phi^+(u,f) = -f'(F^{-1}(\frac{1}{2}+\frac{1}{2}u))/f(F^{-1}(\frac{1}{2}+\frac{1}{2}u))$ corresponds to the most efficient rank test. Putting

$$\chi(x) := \phi^+(2F(|x|)-1)\mathrm{sign}(x) \qquad (3.5)$$

the IC becomes of the form (2.7), so we can apply Hampel's optimality lemma. At the standard normal, the optimally robust solution is given by

$$\tilde{\phi}^+(u) = \min\{\Phi^{-1}(\tfrac{1}{2}+\tfrac{1}{2}u) \;, \; b\} \quad , \qquad (3.6)$$

which is a truncated version of the normal scores function.

4. STATISTICAL PROCEDURES ON TWO SAMPLES

4.1 Definition of the IC

Suppose that for given N we have two samples $x_1,\ldots,x_{m(N)}$ and $y_1,\ldots,y_{n(N)}$. Here $m(N)+n(N)=N$ and $\lim_{N\to\infty} m(N)/N = r$, where $0 < r < 1$. The "location shift" model asserts that the first sample is G-distributed and the second one is F-distributed, where $G(x)=F(x-\theta)$. This θ belongs to Θ ; put $\theta_0=0$. One uses statistics $T_N(x_1,\ldots,x_m;y_1,\ldots,y_n)$ that are invariant with respect to an identical shift of both samples. Assume that T_N converges to the functional $T(H^1,H^2)$ if the observations are iid according to H^1 and H^2 . Define $\xi_N(\theta)$ as the expected value of T_N when $G(x)=F(x-\theta)$, and assume that ξ_N tends to ξ, defined in the same way as the value of T. We say that T is Fisher-consistent if $\xi(\theta)=\theta$ for all θ. Again we suppose that ξ' and ξ^{-1} exist, and we define $U(H^1,H^2):=\xi^{-1}(T(H^1,H^2))$. We now put

$$\begin{aligned} IC_1(T;F;x) &= \lim_{\varepsilon\to 0}[U(F_{\varepsilon,x},F)-U(F,F)]/\varepsilon \\ IC_2(T;F;y) &= \lim_{\varepsilon\to 0}[U(F,F_{\varepsilon,y})-U(F,F)]/\varepsilon \end{aligned} \qquad (4.1)$$

in all points where the limits exist. The case of a parameter of relative scale is treated analogously.

If we suppose the symmetry property $IC_1(T;F;x) = -IC_2(T;F;x)$ to hold, then we can restrict consideration to IC_1 (visualizing the influence of outliers in the first sample). Under the assumptions of [9] it holds that

$$\int IC_1(T;F;x)^2 dF(x) = r(1-r)(c^2)^{-1} \qquad (4.2)$$

where c^2 is the Pitman efficacy. Putting

$$\gamma^*(T,F) = \sup_{x \in \mathbb{R}} |IC_1(T;F;x)| \quad , \qquad (4.3)$$

we have all the elements to apply Hampel's optimality lemma to this case.

4.2 Statistics of the types M and L

Two-sample statistics of the types M and L can be constructed easily. In the location case, one selects a one-sample M-estimator or L-estimator for location which corresponds to a Fisher-consistent functional S and a sequence $\{S_N\}$, and one applies it to both samples. The two-sample statistic

$$T_N(x_1,..,x_m;y_1,..,y_n) = S_m(x_1,..,x_m)-S_n(y_1,..,y_n) \quad (4.4)$$

can then serve as an estimator of the location shift parameter θ or as a test statistic for the null hypothesis $\theta=0$.

In the case of relative scale ($\theta_0=1$), one starts with a one-sample M-estimator or L-estimator for scale, and one replaces subtraction by division in (4.4). In both cases, the resulting T_N converges to a Fisher-consistent functional T, and one obtains

$$IC_1(T;F;x) = IC(S;F;x) = -IC_2(T;F;x) \quad . \qquad (4.5)$$

As we obtain the influence curves of section 2, our previous results can be translated directly to yield optimally robust two-sample statistics of the types M and L.

4.3 Two-sample rank statistics

Two-sample rank statistics for location are given by

$$T_N = \frac{1}{m} \Sigma_{i=1}^{m} a_N(R_i) \tag{4.6}$$

([11]), where R_i is the rank of x_i in the pooled sample. In this subsection, we again assume that F is symmetric. The scores a_N correspond to the scores function $\phi:[0,1]\to\mathbb{R}$ which is nondecreasing and odd (that is, $\phi(1-t)=-\phi(t)$), so $\int\phi(F(x))dF(x) = 0$. The statistics T_N converge to

$$T(H^1,H^2) = \int\phi(rH^1(t)+(1-r)H^2(t))dH^1(t) \quad ,$$

which is not Fisher-consistent. It holds that

$$IC_1(T;F;x) = \phi(F(x))/\int(\phi oF)'(t)dF(t)$$

(see [9]). We can apply Hampel's lemma with

$$\chi(x) = \phi(F(x)) \quad . \tag{4.7}$$

If $\tilde{\chi}$ is skew-symmetric and nondecreasing, then the mapping $\tilde{\phi}(u) = \tilde{\chi}(F^{-1}(u))$ is acceptable and determines an optimally robust two-sample rank statistic. At $F=\Phi$, we obtain

$$\tilde{\phi}(u) = [\Phi^{-1}(u)]_{-b}^{b} \quad , \tag{4.8}$$

a truncation of the Van der Waerden scores function.

Two-sample rank tests for scale are based on (4.6) where the scores correspond to a scores function ϕ_1 , this time supposed to be even ($\phi_1(1-t)=\phi_1(t)$) and nondecreasing for $u \geq \frac{1}{2}$. Because ϕ_1 is defined up to a positive affine transformation, we may still assume that $\int\phi_1(F(t))dF(t)=0$.

It holds that

$$IC_1(T;F;x) = \phi_1(F(x))/\int t(\phi_1 o F)'(t)dF(t)$$

(see [9]). We now put $\chi(x) = \phi_1(F(x))$, obtaining (2.11). If the mapping $\tilde{\chi}$ given by Hampel's lemma is symmetric and nondecreasing for positive arguments, then the scores function $\tilde{\phi}_1(u) = \tilde{\chi}(F^{-1}(u))$ is acceptable and determines an optimally robust solution. At $F=\Phi$, this amounts to

$$\tilde{\phi}_1(u) = [(\Phi^{-1}(u))^2-1-a]_{-b}^{b} \quad , \qquad (4.9)$$

which is a robustification of the Klotz scores function.

The two-sample rank statistics of this section can also be used to derive two-sample estimators of location shift or relative scale (see [9], subsection 3.3). As the resulting estimators possess the same influence curves, one only has to apply the optimally robust scores functions that were determined above.

Acknowledgment The author is much indebted to Prof. F. Hampel for his generous guidance and encouragement.

REFERENCES

[1] HAMPEL, F.: "The Influence Curve and Its Role in Robust Estimation", Journal of the American Statistical Association, 69,pp. 383-393 (1974).

[2] SCHOLZ, F.: "A Comparison of Efficient Location Estimators", Annals of Statistics, 2,pp. 1323-1326 (1974).

[3] BICKEL, P., LEHMANN, E.:"Descriptive Statistics for Nonparametric Models, II. Location", Annals of Statistics, 3,pp. 1045-1069 (1975).

[4] HUBER, P.: "Robust Statistical Procedures", Philadelphia: Society for Industrial and Applied Mathematics (1977).

[5] HUBER, P.: "Robust Estimation of a Location Parameter", Annals of Mathematical Statistics, 35,pp. 73-101 (1964).

[6] JAECKEL, L.: "Robust Estimators of Location: Symmetry and Asymmetric Contamination", Annals of Mathematical Statistics, 42,pp. 1020-1034 (1971).

[7] HUBER, P.: "Robust Statistics: A Review", Annals of Mathematical Statistics, 43,pp. 1041-1067 (1972).

[8] HELMERS, R.: "Edgeworth Expansions for Linear Combinations of Order Statistics", Ph.D. thesis, Leiden, 1978.

[9] ROUSSEEUW, P., RONCHETTI, E.: "Influence Curves of General Statistics", Journal of Computational and Applied Mathematics, 7,pp. 161-166 (1981).

[10] NOETHER, G.: "On a Theorem of Pitman", Annals of Mathematical Statistics, 26,pp. 64-68 (1955).

[11] HAJEK, J., SIDAK, Z.: "Theory of Rank Tests", New York: Academic Press Inc. (1967).