Statistica Sinica **6**(1996), 375-388

ASYMPTOTICS OF AN ESTIMATOR OF A ROBUST SPREAD FUNCTIONAL

Ola Hössjer, Peter J. Rousseeuw and Christophe Croux

Lund Institute of Technology, Universitaire Instelling Antwerpen and Université Libre de Bruxelles

Abstract. The functional $S(F) = \text{med}_{X\sim F}(\text{med}_{Y\sim F}\,|X-Y|)$ proposed by Rousseeuw and Croux (1993) exists for any distribution F, because no moments are needed. It measures spread in the general sense, without reference to a central point. A natural estimator of $S(F)$ is $S_n = \text{med}_i \text{med}_{j,j\neq i}|z_i - z_j|$ where $z_1, \ldots, z_n$ are i.i.d. observations from F. In this paper we prove that S_n is asymptotically normal, and verify that the normality already holds for small samples. Moreover, one can use a constant multiple of S_n to estimate a scale parameter in a (possibly nongaussian) parametric model.

Key words and phrases: Breakdown point, influence function, order statistics, scale estimation.

1. Introduction

Suppose we want to measure the scale of a (population) distribution F on the real line. This can be done with a functional $T(F)$ which is defined over some large family of distributions. The standard deviation functional, although commonly used, is restricted to distributions F with a second moment. In this paper we will focus on more robust functionals.

There are two main classes of scale functionals. The first type (Bickel and Lehmann (1976)) is for symmetric distributions (with arbitrary center of symmetry), where G is said to be more dispersed than F when

$$\Big|G^{-1}(v) - G^{-1}(\frac{1}{2})\Big| \geq \Big|F^{-1}(v) - F^{-1}(\frac{1}{2})\Big| \tag{1.1}$$

for all $0 < v < 1$. They call T a *dispersion functional* if it has the usual invariance properties, and if $T(G) \geq T(F)$ whenever G is more dispersed than F. A typical dispersion functional is the *median absolute deviation* (MAD) given by

$$\text{MAD}(F) = \underset{X\sim F}{\text{med}}\Big|X - F^{-1}(\frac{1}{2})\Big|. \tag{1.2}$$

The MAD was proposed by Hampel (1974) who showed that this functional has a 50% breakdown point, in the sense that it will not tend to zero or infinity when up to half the mass of F is replaced. We can estimate MAD(F) by $\text{MAD}_n := \text{MAD}(F_n)$ where F_n is the empirical distribution of n observations. Apart from being a data-analytic tool in its own right, the MAD_n has also proved very useful in the computation of M-estimators (Andrews et al. (1972), Huber (1981, p. 107)). Further robustness aspects of the MAD can be found in Hampel et al. (1986, Ch. 2).

In a subsequent article, Bickel and Lehmann (1979) considered a more general principle. For arbitrary (not necessarily symmetric) distributions G and F, they say that G is more spread out than F when

$$|G^{-1}(v) - G^{-1}(u)| \geq |F^{-1}(v) - F^{-1}(u)| \tag{1.3}$$

for all $0 < u < v < 1$. They call T a *spread functional* if $T(G) \geq T(F)$ whenever G is more spread out than F. For instance, the interquartile range $T(F) := F^{-1}(\frac{3}{4}) - F^{-1}(\frac{1}{4})$ is a spread functional in the sense of Bickel and Lehmann. Recently, Rousseeuw and Croux (1993) proposed the new spread functional

$$S(F) = \operatorname*{med}_{X\sim F}\Big(\operatorname*{med}_{Y\sim F}|X-Y|\Big), \tag{1.4}$$

where X and Y are independently drawn from F. It uses the distance between X and Y, rather than between X and a central point. The functional S exists for any distribution F because no moment conditions are required, and like the MAD it has a breakdown point of 50%.

A natural estimator of $S(F)$ is given by

$$S_n = \operatorname*{med}_{i}\operatorname*{med}_{j,j\neq i}|z_i - z_j|, \tag{1.5}$$

where $z_1,\ldots,z_n$ are i.i.d. observations from F. This formula is explicit, hence S_n always exists and is unique. Although (1.5) appears to require $O(n^2)$ operations, a faster algorithm has been constructed to compute S_n in only $O(n\log n)$ time (Croux and Rousseeuw (1992)). This also made it possible to study the behavior of S_n empirically, although knowledge of its asymptotic properties remained incomplete.

In this paper the asymptotics of S_n will be studied. This is not a trivial matter because of the "nested" operations in (1.5), which require some additional notation described in Section 2. In Section 3 it is proved that S_n is asymptotically normal, and it is verified empirically that the asymptotics hold already for small samples. The special case of a parametric model with a scale parameter σ is considered in Section 4, where a constant c is computed such that cS_n is a consistent estimator of σ. For the gaussian model, this yields an asymptotic efficiency of 58% (compared to 37% for the MAD).

2. Notation

The functional $S(F)$ in (1.4) can also be written in another way, which will be used in the remainder of the paper. First we formalize the "inner" median. For a fixed z we define

$$L_z(t) = P_F(|z - Z| \leq t) \tag{2.1}$$

and

$$H(z) = L_z^{-1}(0.5) = \operatorname*{med}_{Z \sim F} |z - Z|, \tag{2.2}$$

where $L_z^{-1}(u) = \inf\{t;\, L_z(t) > u\}$ denotes the right continuous inverse. (We use the 0.5-quantile of the distribution function as definition of the median throughout.) For the "outer median" we put

$$L(t) = P_F(H(Z) \leq t) \tag{2.3}$$

so that

$$S(F) = L^{-1}(0.5) = \operatorname*{med}_{Z \sim F} H(Z). \tag{2.4}$$

For proving the asymptotic normality of S_n we impose the following regularity conditions on F, were we write S instead of $S(F)$ for simplicity, and denote by $B(z, r) = [z - r, z + r]$ the closed ball of radius r around z.

(i) There exist $q_1 < q_2$ such that $H(q_1) = H(q_2) = S$.
(ii) There exists $\delta > 0$ such that F has a continuous density f with $0 < \underline{f} < f(z) < \overline{f} < \infty$ on the intervals $B(q_j, \delta)$, $B(q_j - S, \delta)$ and $B(q_j + S, \delta)$ for $j = 1, 2$. Moreover, $\overline{f}_\eta = \sup_{z,y \in B} |f(z) - f(y)|/|z - y|^\eta < \infty$ for some constant η with $0 < \eta < 0.5$, and with $B = B(q_j + S, \delta)$ or $B(q_j - S, \delta)$ for $j = 1$ or 2.
(iii) H is continuous on $B(q_1, \delta)$ and $B(q_2, \delta)$, decreasing on $B(q_1, \delta)$, increasing on $B(q_2, \delta)$, and differentiable at q_1 and q_2 with $H'(q_1) < 0$ and $H'(q_2) > 0$.
(iv) There exist constants $\beta > 0$ and δ' with $0 < \delta' < \delta$ such that

$$\sup_{z \in [q_1 + \delta', q_2 - \delta'[} L_z^{-1}(0.5 + \beta) < S < \inf_{z \notin]q_1 - \delta', q_2 + \delta'[} L_z^{-1}(0.5 - \beta).$$

These conditions seem complicated because they were weakened in order to hold in many situations. (For instance, they hold when F is unimodal with a strictly positive smooth density f.) Condition (ii) is a weaker version of asking that f be Lipshitz and strictly positive, whereas (iv) is a minor regularity condition on the functions L_z. Conditions (i) and (iii) essentially say that the equation $H(z) = S$ has only 2 solutions: one solution q_1 where H is decreasing, and another solution q_2 where H is increasing. It may be possible to generalize this to more than 2 solutions, but this is deferred to future work.

(a)

(b)

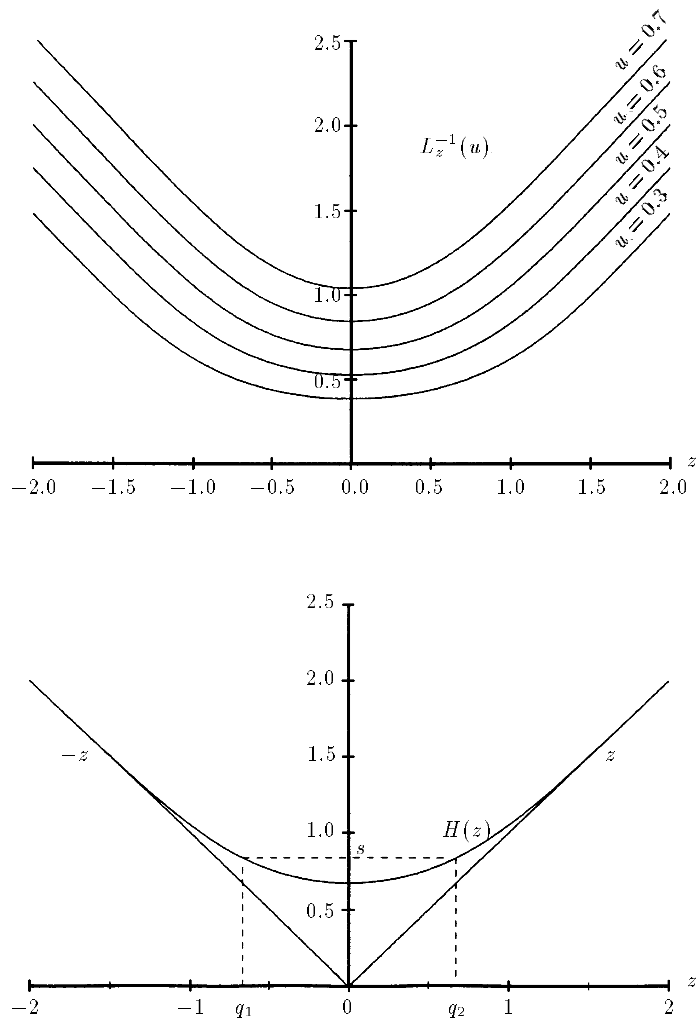

Figure 1. (a) Contours of $L_z(t) = u$ for some values of u; (b) the function $H(z)$.

Let us verify these conditions for F equal to the standard gaussian distribution Φ, which clearly satisfies (ii). Figure 1(a) shows $L_z^{-1}(u)$ for a few values of u. The unimodality and strict positivity of the density $\phi = \Phi'$ imply that $L_z^{-1}(u)$ is strictly increasing in z when $z \geq 0$, and strictly decreasing in z when $z \leq 0$. This makes it easy to verify condition (iv). For computing the function $y(z) = L_z^{-1}(u)$ we could solve

$$\Phi(z + y) - \Phi(z - y) = u \tag{2.5}$$

by the Newton method. A more efficient approach is to solve the differential equation

$$y' = \frac{\phi(z-y)-\phi(z+y)}{\phi(z-y)+\phi(z+y)} \tag{2.6}$$

(obtained by differentiating (2.5) with respect to z) by means of a Runge-Kutta method, with initial condition $y(0)=\Phi^{-1}(\frac{u+1}{2})$. For $z\to\infty$ the functions $L_z^{-1}(u)$ have asymptotes with common slope 1, whereas for $z\to-\infty$ they have slope -1.

The function H of (2.2) corresponds to $u=0.5$. In Figure 1(b) we see that its asymptotes pass through the origin, because, for large $|z|$, $(H(z)-|z|)\to 0$ due to med(Φ)=0. Note that $H^2(z)$ is the functional version of the least median of squares objective function (Rousseeuw and Leroy (1987, pages 164-167)). We see that

$$S = \operatorname*{med}_{Z\sim\Phi} H(Z) = \operatorname*{med}_{Z\sim\Phi} H(|Z|) = H(q_2), \tag{2.7}$$

where $q_2 = \text{med}(|Z|) = \Phi^{-1}(3/4) = -q_1$, yielding $S\approx 0.8385$. This confirms condition (i). For (iii), note that we can compute $H'(q_1)$ and $H'(q_2)$ from (2.6).

3. Asymptotic Normality of S_n

The main result of the paper is the following:

Theorem 3.1. *Under the regularity conditions* (i)–(iv) *on F,*

$$n^{1/2}(S_n-S) = \frac{1}{n^{1/2}}\sum_{i=1}^{n} IF(z_i) + o_p(1) \xrightarrow{d} N\Big(0, \int_{-\infty}^{+\infty} IF(z)^2 dF(z)\Big), \tag{3.1}$$

where S is defined by (1.4) *and*

$$\begin{aligned} IF(z) = \;& \Big(\frac{f(q_2)}{H'(q_2)} - \frac{f(q_1)}{H'(q_1)}\Big)^{-1}\Big(\frac{\text{sgn}(H(z)-S)}{2} + \frac{f(q_2)\text{sgn}(|z-q_2|-S)}{2(f(q_2-S)-f(q_2+S))} \\ & + \frac{f(q_1)\text{sgn}(|z-q_1|-S)}{2(f(q_1+S)-f(q_1-S))}\Big). \end{aligned} \tag{3.2}$$

From (3.1) it follows that the expression in (3.2) is the influence function (in the sense of Hampel (1974)) of the functional $S(F)$. Theorem 3.1 will be proved with three lemmas, in which S_n is approximated by other statistics. Let us write

$$IF(q,z) = \frac{\text{sgn}(|q-z|-H(z))}{2l_z\Big(H(z)\Big)}, \tag{3.3}$$

where $l_z = L'_z$. Hence $IF(\cdot, z_i)$ is the influence function of $\hat{H}(z_i) = \text{med}_{j,j\neq i}|z_j - z_i|$, viewed as an estimator of $H(z_i)$ when z_i is kept fixed. We also use the two quantities

$$\xi_j = \frac{1}{n-1}\sum_{i=1}^{n} IF(z_i, q_j), \quad j=1,2, \tag{3.4}$$

and the random variables

$$V_i = \begin{cases} \xi_1, \; z_i \le q, \\ \xi_2, \; z_i > q, \end{cases} \quad i = 1, \ldots, n,$$

where q is chosen so that $q_1 + \delta < q < q_2 - \delta$ (see (ii) for the definition of δ). We then decompose $\mathbb{R}$ into the regions $A_1 = B(q_1, \delta_n) \cup B(q_2, \delta_n)$ and $A_2 = A_2^- \cup A_2^+ = \{z;\, q_1 + \delta_n < z < q_2 - \delta_n\} \cup \{z;\, z < q_1 - \delta_n \text{ or } z > q_2 + \delta_n\}$ where $\delta_n = n^{-\gamma}$, with $0 < \gamma < 0.5$ a fixed number. Now define the approximating statistics

$$\tilde{S}_1 = \operatorname*{med}_i \{H(z_i) + V_i\} \tag{3.5}$$

and

$$\tilde{S}_2 = \operatorname*{med}_i \{H(z_i) + V_i + W_i\}, \tag{3.6}$$

where

$$W_i = (\hat{H}(z_i) - H(z_i) - V_i) I(z_i \in A_1). \tag{3.7}$$

We first prove that $\tilde{S}_1$ has the asymptotic behaviour claimed for S_n in Theorem 3.1, and then that $\tilde{S}_2$ and S_n are asymptotically equivalent to $\tilde{S}_1$.

Lemma 3.1. *The statistic $\tilde{S}_1$ defined in* (3.5) *satisfies*

$$\tilde{S}_1 = S + \frac{1}{n} \sum_{i=1}^{n} IF(z_i) + o_p(n^{-1/2}). \tag{3.8}$$

Proof. We make use of the "empirical distributions"

$$L_{n0}(z) = \frac{1}{n} \sum_{i=1}^{n} I(H(z_i) \le z) \quad \text{and} \quad L_{n1}(z) = \frac{1}{n} \sum_{i=1}^{n} I(H(z_i) + V_i \le z),$$

so that $\tilde{S}_1 = L_{n1}^{-1}(0.5)$. Denote $\tilde{S}_0 = L_{n0}^{-1}(0.5)$. The difference between L_{n1} and L_{n0} may be written as

$$\begin{aligned} L_{n1}(z) - L_{n0}(z) &= \frac{1}{n} \sum_{z_i \le q} \Big(I(H(z_i) \le z - \xi_1) - I(H(z_i) \le z) \Big) \\ &\quad + \frac{1}{n} \sum_{z_i > q} \Big(I(H(z_i) \le z - \xi_2) - I(H(z_i) \le z) \Big) \\ &= \frac{f(q_1)}{H'(q_1)} \xi_1 - \frac{f(q_2)}{H'(q_2)} \xi_2 + R_1(z - \xi_1, z) + R_2(z - \xi_2, z), \end{aligned} \tag{3.9}$$

where

$$R_1(y, z) = \frac{1}{n} \sum_{z_i \le q} \Big(I(H(z_i) \le y) - I(H(z_i) \le z) \Big) + \frac{f(q_1)}{H'(q_1)} (y - z)$$

and

$$R_2(y,z) = \frac{1}{n}\sum_{z_i>q}\Big(I(H(z_i)\le y) - I(H(z_i)\le z)\Big) - \frac{f(q_2)}{H'(q_2)}(y-z)$$

are regarded as remainder terms when both of their arguments are close to S. Also put $L_{n1}(y) - L_{n1}(z) \triangleq l(S)(y-z) + R_3(y,z)$, where $l(S) = L'(S)$ and R_3 is also considered as a remainder term when both of its arguments are close to S. Also define $u_n = L_{n0}(\tilde{S}_0) - 0.5$ and $v_n = L_{n1}(\tilde{S}_1) - 0.5$, so that $0 \le u_n, v_n \le \frac{1}{2n}$. By the definitions of R_1, R_2 and R_3 we then have

$$\begin{aligned} 0.5 + v_n &= L_{n1}(\tilde{S}_1) = L_{n1}(\tilde{S}_0) + l(S)(\tilde{S}_1 - \tilde{S}_0) + R_3(\tilde{S}_1, \tilde{S}_0) \\ &= 0.5 + u_n + \frac{f(q_1)}{H'(q_1)}\xi_1 - \frac{f(q_2)}{H'(q_2)}\xi_2 + R_1(\tilde{S}_0 - \xi_1, \tilde{S}_0) + R_2(\tilde{S}_0 - \xi_2, \tilde{S}_0) \\ &\quad + l(S)(\tilde{S}_1 - \tilde{S}_0) + R_3(\tilde{S}_1, \tilde{S}_0), \end{aligned}$$

and hence

$$\begin{aligned} \tilde{S}_1 = \tilde{S}_0 &- \frac{f(q_1)}{H'(q_1)l(S)}\xi_1 + \frac{f(q_2)}{H'(q_2)l(S)}\xi_2 + \frac{v_n - u_n}{l(S)} \\ &- \frac{R_1(\tilde{S}_0 - \xi_1, \tilde{S}_0)}{l(S)} - \frac{R_2(\tilde{S}_0 - \xi_2, \tilde{S}_0)}{l(S)} - \frac{R_3(\tilde{S}_1, \tilde{S}_0)}{l(S)}. \end{aligned} \tag{3.10}$$

We observe that

$$\tilde{S}_0 = S + \frac{1}{2l(S)n}\sum_{i=1}^{n}\operatorname{sgn}(H(z_i) - S) + o_p(n^{-1/2}) \tag{3.11}$$

by a standard expansion of the sample median using its influence function. It then follows from (3.2), (3.3), (3.4) and (3.11) that

$$\tilde{S}_0 - \frac{f(q_1)}{H'(q_1)l(S)}\xi_1 + \frac{f(q_2)}{H'(q_2)l(S)}\xi_2 = S + \frac{1}{n}\sum_{i=1}^{n} IF(z_i) + o_p(n^{-1/2}). \tag{3.12}$$

Comparing (3.10) with (3.12), we see that it suffices to show that the remainder terms in (3.10) are all $o_p(n^{-1/2})$. To begin with, $(|v_n - u_n|)/l(S) \le 1/(2l(S)n)$. In order to handle R_1 to R_3 we first note that $\tilde{S}_0 - S, \xi_1$ and ξ_2 are all $O_p(n^{-1/2})$ in view of (3.4) and (3.11). Since $|\tilde{S}_1 - S| \le |\tilde{S}_0 - S| + \max(|\xi_1|, |\xi_2|)$, we also have that $\tilde{S}_1 - S$ is $O_p(n^{-1/2})$. Hence, it suffices to show that for any $A > 0$,

$$\sup_{z,y\in B(S,A/\sqrt{n})} |R_j(z,y)| = o_p(n^{-1/2}), \quad j = 1,2,3. \tag{3.13}$$

Consider first $j=1$ in (3.13). Define the function $H^*(z) = H(z)I(z \le q) + (S + 2A)I(z > q)$, and put $L^*(z) = P_F(H^*(Z) \le z)$ for the corresponding distribution function. Then

$$L_n^*(z) = \frac{1}{n}\sum_{i=1}^{n} I(H^*(z_i) \le z) = \frac{1}{n}\sum_{z_i \le q} I(H(z_i) \le z), \quad \forall z \le S + A. \tag{3.14}$$

It is not hard to see that the regularity conditions on F imply that $L^*(z)$ is differentiable at $z = S$, with $l^*(S) = L^{*\prime}(S) = -f(q_1)/H'(q_1)$. Consequently, by (3.9) and (3.14),

$$\begin{aligned}\sup_{z,y\in B(S,A/\sqrt{n})} |R_1(z,y)| \le &\sup_{z,y\in B(S,A/\sqrt{n})} |L_n^*(z) - L_n^*(y) - L^*(z) + L^*(y)| \\ &+ \sup_{z,y\in B(S,A/\sqrt{n})} |L^*(z) - L^*(y) - l^*(S)(z-y)|.\end{aligned} \tag{3.15}$$

The first term on the RHS of (3.15) concerns the oscillation of the empirical distribution $L_n^*(z)$ locally around $z = S$, and it is $O_p(n^{-3/4}(\log n)^{1/2})$ by Lemma 6.3.2 in Reiss (1988). The second term on the RHS of (3.15) is $o(n^{-1/2})$ since $L^*(z)$ is differentiable at $z = S$. Hence we have proved (3.13) for $j = 1$, and the case $j = 2$ is treated in the same way. Finally, $R_3(y,z) = R_1(y - \xi_1, z - \xi_1) + R_2(y - \xi_2, z - \xi_2)$, so (3.13) also follows for $j = 3$, since we may pick any (other) value of A in (3.13) for $j = 1, 2$, and use the fact that ξ_1 and ξ_2 are both $O_p(n^{-1/2})$.

Lemma 3.2. *The statistics $\tilde{S}_1$ and $\tilde{S}_2$ defined in* (3.5)–(3.6) *are asymptotically equivalent, that is*

$$\tilde{S}_2 - \tilde{S}_1 = o_p(n^{-1/2}). \tag{3.16}$$

Proof. In view of (3.3), we make the expansion

$$\hat{H}(z_i) = H(z_i) + \frac{1}{n-1}\sum_{j,j\ne i} IF(z_j, z_i) + R_i \triangleq H(z_i) + T_i + R_i.$$

Clearly,

$$|\tilde{S}_2 - \tilde{S}_1| \le \max_{z_i \in B(q_1,\delta_n)} |T_i - \xi_1| + \max_{z_i\in B(q_2,\delta_n)} |T_i - \xi_2| + \max_{z_i \in A_1} |R_i|. \tag{3.17}$$

The lemma is established by verifying that each of the three terms on the RHS of (3.17) is $o_p(n^{-1/2})$. Starting with the first term,

$$T_i - \xi_1 = -\frac{IF(z_i, q_1)}{n-1} + \frac{1}{n-1}\sum_{j,j\ne i}\Big(IF(z_j, z_i) - IF(z_j, q_1)\Big),$$

and hence

$$\max_{z_i \in B(q_1,\delta_n)} |T_i - \xi_1| \leq \frac{1}{2l_{q_1}(S)(n-1)} + \max_{z_i \in B(q_1,\delta_n)} |\Delta(z_i)|,$$

where $\Delta(z_i) = I(|z_i - q_1| \leq \delta_n) \sum_{j, j\neq i} Y_{ij}/(n-1)$, and $Y_{ij} = IF(z_i, z_j) - IF(q_1, z_j)$. In order to apply a large deviation inequality on $\Delta(z_i)$ (and $\max_{z_i \in B(q_1,\delta_n)} |\Delta(z_i)|$) we need some preparations. First $P(|Y_{ij}| \leq M) = 1$, where

$$M = \frac{1}{2l_{q_1}(S)} + \sup_{z \in B(q_1,\delta_n)} \frac{1}{2l_z(H(z))}.$$

In order to show that $M < \infty$, we need to establish that

$$\inf_{z \in B(q_1,\delta_n)} l_z(H(z)) \geq \underline{l} > 0, \tag{3.18}$$

for some positive constant $\underline{l}$. Since

$$l_z(H(z)) = f(z + H(z)) - f(z - H(z)), \tag{3.19}$$

H is continuous at q_1 by (iii), and f is lower bounded away from 0 in neighborhoods of $q_1 \pm S$ according to (ii), it follows that (3.18) holds for large enough n (i.e., small enough δ_n). We find that

$$E(Y_{ij}^2 | z_i) \leq \frac{2}{l_{q_1}(S)} P_F\Big(Z \in B_{z_i} \triangle B_{q_1}\Big) + \frac{1}{2\underline{l}^4}\Big(l_{z_i}(H(z_i)) - l_{q_1}(S)\Big)^2,$$

where $B_z = \{z'; |z' - z| > H(z)\}$ and $\underline{l}$ is defined in (3.18). Hence,

$$\begin{aligned} B_z \triangle B_{q_1} &= [q_1 - S, q_1 + S] \triangle [z - H(z), z + H(z)] \\ &\subseteq [q_1 - S \pm (|z - q_1| + |H(z) - S|)] \cup [q_1 + S \pm (|z - q_1| + |H(z) - S|)] \\ &\subseteq [q_1 - S \pm C|z - q_1|] \cup [q_1 + S \pm C|z - q_1|], \end{aligned} \tag{3.20}$$

where the last inequality holds for $|z - q_1|$ small enough, using the fact that H is differentiable at q_1 (and hence we may choose $C = 1 + 2|H'(q_1)|$ for instance). Assume now that n is so large that (3.20) holds for all $z \in B(q_1, \delta_n)$ and moreover that $C|z - q_1| \leq \delta$ (cf. (ii)). Then

$$P_F\Big(Z \in B_z \triangle B_{q_1}\Big) \leq 4C\overline{f}|z - q_1|, \quad \forall z \in B(q_1, \delta_n). \tag{3.21}$$

Next, because of (3.19),

$$\begin{aligned} |l_z(H(z)) - l_{q_1}(S)| &\leq |f(z + H(z)) - f(q_1 + S)| + |f(z - H(z)) - f(q_1 - S)| \\ &\leq 2\overline{f}_\eta\Big(|z - q_1| + |H(z) - S|\Big)^\eta \leq 2\overline{f}_\eta C^\eta |z - q_1|^\eta, \end{aligned} \tag{3.22}$$

with C the same constant as in (3.20) and $\overline{f}_\eta$ as defined in (ii). It follows from (3.21)–(3.22) that

$$\max_{z_i \in B(q_1,\delta_n)} \max_{j;j\neq i} E(Y_{ij}^2|z_i) \leq C'\delta_n^{2\eta} \tag{3.23}$$

for some constant C'. Because of (3.23) and the uniform boundedness of Y_{ij}, we may apply an exponential inequality due to Bernstein (cf. Pollard (1984, Appendix B)) to estimate tail probabilities of each $T_i - \xi_1$, $z_i \in B(q_1, \delta_n)$. This gives

$$\max_{z_i \in B(q_1,\delta_n)} |T_i - \xi_1| = O_p\Big(\frac{\delta_n^\eta \log n}{n^{1/2}}\Big) = O_p(n^{-\gamma\eta - 1/2}\log n) = o_p(n^{-1/2}).$$

The second term on the RHS of (3.17) is treated in the same way as the first. It remains to consider the last term, containing the Bahadur representation remainder terms. Let $U_\varepsilon = [0.5-\varepsilon, 0.5+\varepsilon]$. Provided we find an $\varepsilon > 0$ such that for $n \geq n_0(\varepsilon)$,

$$\underline{l}_\varepsilon = \inf_{(z,u)\in A_1\times U_\varepsilon} |l_z(L_z^{-1}(u))| > 0,$$

and

$$\overline{L}_\varepsilon = \sup_{\substack{(z,u)\in A_1\times U_\varepsilon \\ u\neq 0.5}} \frac{|l_z(L_z^{-1}(u)) - l_z(H(z))|}{|L_z^{-1}(u) - H(z)|^\eta} < \infty,$$

it follows that the Bahadur remainder terms are uniformly small in the sense that

$$\max_{z_i\in A_1} |R_i| = O_p\Big((n^{-1}\log n)^{(1+\eta)/2}\Big) = o_p(n^{-1/2}). \tag{3.24}$$

We observe that $l_z(L_z^{-1}(u)) = f(z+L_z^{-1}(u)) + f(z-L_z^{-1}(u))$, and since f is lower bounded away from 0 on $B(q_j - S, \delta)$ and $B(q_j + S, \delta)$ for $j = 1, 2$ (cf. (ii)), we may choose ε so small and $n_0(\varepsilon)$ so large that

$$z \pm L_z^{-1}(u) \in B(q_1 - S,\delta)\cup B(q_1+S,\delta)\cup B(q_2-S,\delta)\cup B(q_2+S,\delta) \tag{3.25}$$

whenever $(z,u) \in A_1 \times U_\varepsilon$. But (3.25) implies that $\underline{l}_\varepsilon \geq 2\underline{f} > 0$ with $\underline{f}$ as defined in (ii), and $\overline{L}_\varepsilon \leq 2\overline{f}_\eta < \infty$, which proves (3.24).

Lemma 3.3. *The statistics $\tilde{S}_2$ and S_n defined in* (3.6) *and* (1.5) *are asymptotically equivalent, that is*

$$S_n - \tilde{S}_2 = o_p(n^{-1/2}). \tag{3.26}$$

Proof. Let $H_n = \min(H(q_1-\delta_n)-S, S-H(q_1+\delta_n), S-H(q_2-\delta_n), H(q_2+\delta_n)-S)$. The fact that H is strictly monotone on $B(q_j,\delta)$ for $j = 1,2$ and (iv) imply that for large enough n, $H(z) \leq S - H_n$ when $z \in A_2^-$ and $H(z) \geq S + H_n$ when $z \in A_2^+$. Hence

$$P(S_n \neq \tilde{S}_2) \leq P(|\tilde{S}_2 - S| \geq (1-\rho)H_n) + P(\max_{z_i \in A_2} |V_i| \geq \rho H_n) \tag{3.27}$$
$$+ P(\max_{z_i \in A_2^-} \hat{H}(z_i) \geq S-(1-\rho)H_n) + P(\min_{z_i \in A_2^+} \hat{H}(z_i) \leq S+(1-\rho)H_n),$$

with $0 < \rho < 1$ an arbitrary number. It remains to show that the RHS of (3.27) tends to 0 as $n \to \infty$. Since H is differentiable at q_1 and q_2 with nonzero derivatives, it is clear that

$$C_1\delta_n \leq H_n \leq C_2\delta_n \tag{3.28}$$

for some positive constants C_1 and C_2. We also know from Lemma 3.1–3.2 that $\tilde{S}_2 - S = O_p(n^{-1/2})$ and $\max_{z_i \in A_2} |V_i| \leq \max(|\xi_1|, |\xi_2|) = O_p(n^{-1/2})$. Therefore, the first two terms on the RHS of (3.27) tend to zero as $n \to \infty$. Since the last two terms are treated in the same way, we confine ourselves to a study of the third. Let $\underline{f}$ be as in (ii) and pick ρ' so that $0 < \rho' < 2\underline{f}\rho$. Then we will prove

$$\sup_{z \in A_2^-} L_z^{-1}(0.5 + \rho' H_n) \leq S - (1-\rho)H_n. \tag{3.29}$$

In view of (iv) and (3.28), it is clear that (3.29) must hold for all $z \in [q_1+\delta', q_2-\delta']$ simultaneously, provided n is large enough (how large n must be chosen depends on ρ and ρ'). Suppose now that $z \in [q_1 + \delta_n, q_1 + \delta'] \cup [q_2 - \delta', q_2 - \delta_n]$. Then

$$\begin{aligned} &P_F(|Z-z| \leq S - (1-\rho)H_n) \\ &= P_F(|Z-z| \leq S - H_n) + P_F(S - H_n < |Z-z| \leq S - (1-\rho)H_n) \\ &\geq 0.5 + 2\underline{f}\rho H_n \geq 0.5 + \rho' H_n, \end{aligned}$$

since for large enough n the set $\{z'; S - H_n < |z'-z| \leq S - (1-\rho)H_n\}$ is contained in the union of the four closed balls $B(q_j \pm S, \delta)$, $j = 1, 2$ on which f is lower bounded by $\underline{f}$. For ease of notation, put $\overline{H}(z) = L_z^{-1}(0.5 + \rho' H_n)$, and let $L_{n-1,z_i}(y) = \sum_{j, j\neq i} I(|z_j - z_i| \leq y)/(n-1)$ denote the empirical distribution corresponding to L_{z_i}. Our next objective is to show that

$$\max_{z_i \in A_2^-} \Big| L_{n-1,z_i}(\overline{H}(z_i)) - L_{z_i}(\overline{H}(z_i)) \Big| = O_p\Big((\log n)^{1/2} n^{-1/2}\Big). \tag{3.30}$$

However, (3.30) is a consequence of Hoeffding's exponential inequality (see Pollard (1984, Appendix B)), which in our case implies (after first conditioning on z_i) that

$$P\Big(|L_{n-1,z_i}(\overline{H}(z_i)) - L_{z_i}(\overline{H}(z_i))| \geq t(n-1)^{-1/2}\Big) \leq 2\exp(-2t^2).$$

Since $L_{z_i}(\overline{H}(z_i)) \geq 0.5 + \rho' H_n$ by our definition of inverse distribution functions, it follows from (3.28) and (3.29) that with probability tending to one, $\min_{z_i \in A_2^-} L_{n-1,z_i}(\overline{H}(z_i)) > 0.5$, and hence because of (3.29), $\max_{z_i \in A_2^-} \hat{H}(z_i) \leq \max_{z_i \in A_2^-} \overline{H}(z_i) \leq S - (1-\rho)H_n$ with probability tending to one. This shows that the third term on the RHS of (3.27) also goes to 0 as $n \to \infty$.

Proof of Theorem 3.1. Writing S_n as a telescoping sum, $S_n = \tilde{S}_1 + (\tilde{S}_2 - \tilde{S}_1) + (S_n - \tilde{S}_2)$, Theorem 3.1 now follows by applying Lemmas 3.1–3.3 to the three terms in this sum, using the Central Limit Theorem and Slutsky's Lemma.

Table 1. Finite-sample behavior of S_n

n	average(S_n)	n(var(S_n))
10	.832	.78
20	.838	.69
40	.838	.62
60	.839	.63
80	.840	.62
100	.836	.61
200	.839	.61
∞	.839	.60

In order to verify how well the asymptotic result in Theorem 3.1 applies to the finite-sample behavior of S_n we carried out some numerical experiments. By means of Q-Q plots we checked that the sampling distribution of S_n is approximately gaussian. Moreover, it is important to know how large n has to be before the asymptotic variance reflects the finite-sample situation. For each value of n in Table 1 we generated 10,000 gaussian samples of that size. The first column lists the average estimated value, which is very close to the limit $S(\Phi) = 0.8385$ obtained in Section 2. The other column of Table 1 contains n times the empirical variance of S_n. The asymptotic variance was computed from (3.1), yielding $\int IF^2(z)d\Phi(z) = 0.6028$. We see that the finite-sample variance of S_n converges rapidly to its limit, the difference being small already at $n = 40$.

4. Estimation in a Parametric Model

In the preceding sections we have considered the functional S as a nonparametric measure of spread (in the sense of Bickel and Lehmann (1979)) which can be computed for any distribution. However, let us now suppose that we are in the particular case of a parametric model with a scale parameter σ, which may be a pure scale model $F_\sigma(z) = F(z/\sigma)$ for $0 < \sigma < \infty$, or a location-scale model

$F_{\mu,\sigma}(z) = F((z-\mu)/\sigma)$ for $-\infty < \mu < \infty,\ 0 < \sigma < \infty$. In either model we can use the functional S to estimate σ, because S is equivariant under multiplication by a scalar and invariant under translation. In practice we will use the functional $T(G) = cS(G)$ where c is a positive constant such that T is Fisher consistent for σ, meaning that $T(F_\sigma) = \sigma$ (or $T(F_{\mu,\sigma}) = \sigma$ in the location-scale model). For this it suffices to set $c := 1/S(F)$. By Theorem 3.1 the estimator $cS_n(z_1, \ldots, z_n)$ is then consistent for σ in the usual sense.

For the gaussian model we find $c = 1/S(\Phi) = 1/0.8385 = 1.1926$ according to Section 2. For other distributions F we can compute $S(F)$ in an analogous way. Table 2 gives the formula of H and the corresponding value $S(F)$ for the triangular, Laplace, logistic and Cauchy distributions. In the latter case $H(z) = \sqrt{z^2+1}$ is a hyperbola, which looks similar to Figure 1(b). Table 2 also contains an asymmetric distribution (the negative exponential) and a bimodal mixture of two Cauchy distributions. For distributions for which H cannot be written down analytically (such as the gaussian distribution), we solve (2.5) or (2.6) numerically to obtain $S(F)$.

In a parametric model one can compare the variability of an estimator with the Cramér-Rao bound. For the gaussian model, Theorem 3.1 implies that 1.1926 S_n has an asymptotic efficiency of 58% (whereas the equally robust estimator 1.4826 MAD_n yields 37%). Similar computations are possible in a scale model based on an asymmetric model distribution. In the negative exponential model, the estimator cS_n (where $c = 1/0.5888$ from Table 2) has 55% efficiency. For heavy tailed models the efficiency of robust estimators is higher, and in fact for the Cauchy model (where $c = 1/\sqrt{2}$) we obtain 95%.

Table 2. The value $S(F)$ for several distributions F

F	$f(z)$	$H(z)$	$S(F)$
Triangular	$\max(1-\lvert z\rvert, 0)$	$\begin{cases} 1-\sqrt{1/2-z^2} & \lvert z\rvert \le 1/2 \\ \lvert z\rvert & \lvert z\rvert \ge 1/2 \end{cases}$	$1-\sqrt{\sqrt{2}-1}$
Laplace	$\frac{1}{2}\exp(-\lvert z\rvert)$	$\ln(2\cosh(z))$	$\ln(\frac{5}{2})$
Logistic	$\frac{e^z}{(1+e^z)^2}$	$2\ \mathrm{arctanh}(\frac{1+\cosh z}{2+\sqrt{3+\cosh^2 z}})$	$2\ \mathrm{arctanh}(\sqrt{13}-3)$
Cauchy	$\frac{1}{\pi}\ \frac{1}{1+z^2}$	$\sqrt{z^2+1}$	$\sqrt{2}$
Negative exponential	$\exp(-z)1_{[0,\infty[}(z)$	$\begin{cases} \ln(2)-z & z \le \ln(2)/2 \\ \mathrm{arcsinh}(e^z/4) & z \ge \ln(2)/2 \end{cases}$	0.5888
Bimodal Cauchy	$\frac{1}{2\pi}\{(1+(z-m)^2)^{-1}+ (1+(z+m)^2)^{-1}\}$	$\sqrt{z^2+m^2+1}$	$\sqrt{2(1+m^2)}$

Acknowledgements

The research of the first author was supported by the Swedish Board for Technical Development under contract 88-01359P. The third author was supported by the Belgian National Fund for Scientific Research. We would like to thank the Editor, Associate Editor, and two referees for helpful comments.

References

Andrews, D. F., Bickel P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). Robust Estimates of Location: Survey and Advances. Princeton University Press, Princeton NJ.

Bickel, P. J. and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models III: Dispersion. Ann. Statist. **4**, 1139-1158.

Bickel, P. J. and Lehmann, E. L. (1979). Descriptive statistics for nonparametric models IV: Spread. In Contributions to Statistics, Hájek Memorial Volume (Edited by J. Jurečková), 33-40. Academia, Prague.

Croux, C. and Rousseeuw, P. J. (1992). Time-efficient algorithms for two highly robust estimators of scale. In Computational Statistics, Volume 1 (Edited by Y. Dodge and J. Whittaker), 411-428, Physica-Verlag, Heidelberg.

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

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

Huber, P. J. (1981). Robust Statistics. John Wiley, New York.

Pollard, D. (1984). Convergence of Stochastic Processes. Springer-Verlag, New York.

Reiss, R.-D. (1989). Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics. Springer-Verlag, New York.

Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. J. Amer. Statist. Assoc. **88**, 1273-1283.

Rousseeuw, P. J. and Leroy, A. M. (1987). Robust Regression and Outlier Detection. John Wiley, New York.

Department of Mathematics and Statistics, Lund Institute of Technology, Box 118, S-22100 Lund, Sweden.

Department of Mathematics and Computing, Universitaire Instelling Antwerpen, Universiteitsplein 1, B-2610 Antwerp, Belgium.

Department CEME-CP 139, Université Libre de Bruxelles, Avenue Roosevelt 150, B-1050 Brussels, Belgium.

(Received April 1993; accepted April 1995)