

On locally uniformly linearizable high breakdown
location and scale functionals

P.L.Davies

April 16, 1998

1 Contents

Section 2 contains a discussion of the reasons for constructing locally uniformly linearizable high breakdown functionals. The definition of locally uniform linearizability is given in Section 3. A first construction based on M-functionals is given in Section 4 which is used in Section 5 to give a locally uniformly linearizable weighted mean. A second construction is given in Section 6. Finally in Section 7 we apply the weighted mean based on the M-functional to some models and real data sets and compare its performance with those of some non-linearizable functionals.

2 Introduction

One of the simplest statistical problems is the location-scale problem on the real line. Given a data set $\mathbf{x}_n = \{x_1, \dots, x_n\}$ we are required to specify two numbers L and S , together with upper and lower bounds, which describe the location and the scale respectively of the data. In spite of its apparent simplicity the problem has as yet no satisfactory solution. Most approaches including robust ones are based on a central model F_0 which is assumed to be true or to contain the truth within some small metric ball. An exception is Morgenthaler and Tukey [19] where k different models or challenges are considered simultaneously. Data rarely come accompanied by a central model and when analysing large numbers of data sets in an automatic manner such an approach is unwarranted. One possibility is to use non-parametric measures of location and scale as in Bickel and Lehmann [5] [6] [7]. This is the approach we shall adopt but with the added restriction of requiring stability of analysis over a wide range of models and data sets. By stability of analysis we mean stability of quantiles of the sampling distribution. This is a much stronger demand than the continuity of the functional itself as considered in Section 4 of Bickel and Lehmann [6].

By location and scale functionals we mean functionals which have the appropriate equivariance properties with respect to affine transformations. We shall denote a generic location functional by T_L and a generic scale functional by T_S . We shall limit the discussion to the location functional, treating the scale as a nuisance parameter. For data \mathbf{x}_n the point value of the location functional is $T_L(P_n(\mathbf{x}_n))$ where $P_n(\mathbf{x}_n)$ denotes the empirical distribution associated with the data. The set of acceptable values of the location functional is given by

$$A(\mathbf{x}_n, T_L) = \{T_L(P) : P \text{ is an adequate approximation for the data}\}$$

(see Davies [12]) where we use the following two concepts of approximation. The first is based on the Kuiper metric defined by

$$d_{ku}(P, Q) = \sup\{|P(I) - Q(I)| : I \text{ an interval}\}.$$

In the sense of the Kuiper metric a distribution P will be considered to be an adequate approximation if

$$d_{ku}(P_n(\mathbf{x}_n), P) \leq qu(n, 0.99, d_{ku}, P) \quad (1)$$

where $qu(n, \alpha, d_{ku}, P)$ denotes the α -quantile of the Kuiper metric. Asymptotic values can be derived from Proposition 12.3.6 of Dudley [13].

The second concept requires that the value of the location functional T_L for the real data is close to size n deriving from the distribution P . We base this concept of approximation on the studentized location functional T_{LS} defined by

$$T_{LS}(Q, P) = (T_L(Q) - T_L(P)) / T_S(Q).$$

In the sense of T_{LS} a distribution P will be considered to be an adequate approximation if

$$\begin{aligned} qu(n, 0.02, T_{LS}, P) / \sqrt{n} &\leq T_{LS}(P_n(\mathbf{x}_n), P) \\ &\leq qu(n, 0.98, T_{LS}, P) / \sqrt{n} \end{aligned}$$

where $qu(n, \alpha, T_{LS}, P)$ denotes the α -quantile of the statistic $\sqrt{n}T_{LS}(P_n(P), P)$ under the model P . Here $P_n(P)$ denotes the empirical distribution of n i.i.d. random variables with distribution P . We note that if the data \mathbf{x}_n really are the realization of i.i.d. random variables with distribution P , then P will be regarded as an adequate approximation with a probability of The approximation region for the values of T_L based on both concepts of approximation is given by

$$\begin{aligned} A(\mathbf{x}_n, T_L) = \{ &T_L(P) : d_{ku}(P_n(\mathbf{x}_n), P) \leq qu(n, 0.99, d_{ku}, P), \\ &qu(n, 0.02, T_{LS}, P) / \sqrt{n} \leq T_{LS}(P_n(\mathbf{x}_n), P) \\ &\leq qu(n, 0.98, T_{LS}, P) / \sqrt{n}\}. \end{aligned} \quad (2)$$

The discussion so far applies to any functionals T_L and T_S but the formulation of (2) requires functionals which are well-defined for a large class of distributions. This is one of the demands we shall make. Secondly we wish to allow for the possibility that the data include a proportion of erroneous values. Even in this situation we do not want the functionals to fail completely and to avoid this we shall require a high local breakdown point. Thirdly it is often important for applications that the functional be interpretable. In my experience many practitioners are unhappy with M-functionals but will accept weighted means if the weights seem reasonable. Two further reasons for considering weighted means are that they allow a smooth adaption to a finite set of different models (Section 6) and that they give an additional flexibility which can be important, for example if the scale functional is of independent interest (Croux [10]). Fourthly we shall be interested in functionals for which the statistical analysis is stable. This will be so if the values $T_L(P)$, $T_S(P)$ and $qu(n, \alpha, T_{LS}, P)$ are continuous as a function of P with respect to the Kuiper metric. This is not the case for the median nor for the mean for which it is known that there does not

exist any non-trivial non-parametric confidence interval, Bahadur and Savage [2]. It is the case for functionals T_L and T_S which are locally uniformly linearizable in the sense to be defined below. For such functionals the approximation region itself can be approximated by

$$\begin{aligned}
A(\mathbf{x}_n, T_{LS}) = & \\
& [T_L(P_n(\mathbf{x}_n)) - qu(n, 0.98, T_{LS}, P_n(\mathbf{x}_n))T_S(P_n(\mathbf{x}_n))/\sqrt{n}, \\
& T_L(P_n(\mathbf{x}_n)) - qu(n, 0.02, T_{LS}, P_n(\mathbf{x}_n))T_S(P_n(\mathbf{x}_n))/\sqrt{n}] \quad (3)
\end{aligned}$$

which is nothing more than the bootstrap estimate.

A further consequence of locally uniform linearity is that the functionals are asymptotically normal. If we denote the influence function of T_L at the point P by $I(\cdot, T_L, P)$ then we have

$$\sqrt{n}(T_L(P_n(P)) - T_L(P)) \Rightarrow N(0, \sigma(T_L, P)^2)$$

where

$$\sigma(T_L, P)^2 = \int I(u, T_L, P)^2 dP(u).$$

We shall require that $\sigma(T_L, P)^2$ is continuous in a Kuiper neighbourhood and consequently we may replace P by $P_n(\mathbf{x}_n)$ to give the following approximation to the approximation interval;

$$\begin{aligned}
A(\mathbf{x}_n, T_{LS}) = & \quad [T_L(P_n(\mathbf{x}_n)) - 2.05\sigma(T_L, P_n(\mathbf{x}_n))/\sqrt{n}, \\
& T_L(P_n(\mathbf{x}_n)) + 2.05\sigma(T_L, P_n(\mathbf{x}_n))/\sqrt{n}] \quad (4)
\end{aligned}$$

which is the asymptotic bootstrap approximation.

Huber [18] page 5 has a list of desirable features which a functional should exhibit. These include reasonably good efficiency at the assumed model and stable asymptotic behaviour in a neighbourhood of the model. Bickel and Lehmann [6] consider the relative efficiencies of different non-parametric estimators at certain specified models. As we are also concerned with a situation where there is no central model pointwise efficiency considerations, even over neighbourhoods of a single model, are of limited relevance. What is important is the performance over a large range of different models or challenges (Morgenthaler and Tukey [19]) and real data sets. The measures of performance may include the number of complete failures, the lengths of the approximation intervals as well as the accuracy of these intervals and the computational costs. Putting all these requirements together we have the following list of desirable properties:

DP1 Affine equivariance

DP2 A large domain

DP3 A high local breakdown point

DP4 Interpretability of the functional.

DP5 Locally uniform differentiability

DP6 Superior performance over existing competitors

Morgenthaler and Tukey [19] consider an approach which in one sense lies between the non-parametric approach of this paper and the standard one model situation of robust statistics. They consider a finite number of models or challenges and look for a procedure which performs well at all of them. The hope is that such a procedure will also perform reasonably well for challenges which lie between. For a given sample a likelihood based compromise between the two challenges is made. The use of likelihood means that the method of Morgenthaler and Tukey does not satisfy DP5. In Section 6 we show how it is possible to "coarsen" a large class of distributions by reducing them to a finite sample of m points which themselves satisfy DP5. These points can be used to decide between a finite set of challenges and hence to make the weights of the weighted mean depend on the shape of the sample but in a differentiable manner.

3 Local uniform linearity

3.1 Notation and definition

To fulfill DP2 the domains we consider are of the form

$$W(\eta) = \{P : \Delta(P) < \eta\}$$

where $\Delta(P)$ denotes the size of the largest atom of P . The Kuiper ball of radius δ and with centre P will be denoted by $B(P, \delta)$. Given any P in $W(\eta)$ it is clear that there exists a $\delta(P)$ such that $B(P, \delta(P)) \subset W(\eta)$. The local breakdown point $\epsilon^*(P, T)$ of T at P is defined by

$$\epsilon^*(P, T) = \sup\{\epsilon > 0 : \sup_{Q \in B(P, \epsilon)} (|T_L(Q)| + |\log T_S(Q)|) < \infty\} \quad (5)$$

The variation of a function $h : \mathbb{R} \rightarrow \mathbb{R}$ will be denoted by $\|h\|_v$: the variation $\|H\|_v$ of a function $H : \mathbb{R} \rightarrow \mathbb{R}^2$ is defined to be the sum of the variations of the components. We use the usual order of magnitude notation but shall write O_P to emphasize that the constant involved depends only on the distribution P .

The functional $T : W(\eta) \rightarrow \mathbb{R}^2$ is said to be locally uniformly linearizable if the following hold:

L1 There exists a function $I(\cdot, T, \cdot) : \mathbb{R} \times W(\eta) \rightarrow \mathbb{R}^2$ with the following properties for each $P \in W(\eta)$:

L1.1 $I(\cdot, T, P) : \mathbb{R} \rightarrow \mathbb{R}^2$ is of bounded variation

L1.2 There exists a $\delta(P) > 0$ such that

$$\sup_u \|I(u, T, Q) - I(u, T, P)\| = O_P(d_{ku}(Q, P))$$

for all $Q \in B(P, \delta(P))$

L1.3 $\int I(u, T, P)dP(u) = 0$

L1.4 $\int I_j(u, T, P)^2 dP(u) > 0$ for each component j

L2 There exists a strictly increasing function $\tau : \mathbb{R} \rightarrow \mathbb{R}$ satisfying

$\lim_{u \rightarrow 0} \tau(u)/u = 0$ such that the following holds. There exists a $\delta(P) > 0$ such that for all Q and Q' in $B(P, \delta(P))$

$$\|T(Q') - T(Q) - \int I(u, T, Q)dQ'(u)\| = O_P(\tau(d_{ku}(Q', Q))).$$

The function $I(\cdot, T, P)$ is the influence function of the functional T . The requirements here are stronger than those of Hampel, Ronchetti, Rousseeuw and Stahel [16] and of Rieder [20]. Hampel et al [16] consider fixed P and deviations Q of the form $(1 - \epsilon)P + \epsilon\delta_x$ where δ_x denotes the unit mass in x . Rieder also restricts attention to fixed P but with Q and Q' in a shrinking neighbourhood of P . The regularity conditions placed on $I(\cdot, T, P)$ are also stronger than those required by Hampel et al [16] and Rieder [20].

3.2 The asymptotics of locally uniformly linearizable functionals

In this section we show that locally uniformly linearizable functionals converge locally uniformly to a normal distribution. A referee pointed out that the location functional without an auxiliary scale functional which is based on Huber's minimax estimator (Huber [17]) $\psi(x) = (-k) \vee x \wedge k$ has uniform asymptotics for a restricted class of distributions, namely for all symmetric distributions in the ϵ -contamination ball of the standard normal distribution with $F\{\pm k\} = 0$. Rieder [20] uses uniformity in a local asymptotic sense to derive asymptotic minimax and convolution theorems for the estimation of functionals. Bednarski, Clarke and Kolkiewicz [4] consider the relationship between asymptotic expansions and Fréchet differentiability. Most results on differentiability are pointwise but such results do not guarantee robustness or the success of bootstrapping for any sample. The proofs mimic those of Clarke [9], Huber [17], Scholz [21], Serfling [22] and Shao [23] amongst others.

The formula we use for partial integration is

$$d(FG) = \tilde{F}dG + \tilde{G}dF$$

where

$$\tilde{h}(u) = \frac{h(u+0) + h(u-0)}{2} \tag{6}$$

for any non-decreasing function h . We have

Theorem 3.1

Let $h : \mathbb{R} \rightarrow \mathbb{R}$ be of bounded variation and F and G be two distribution functions. Then

$$\left| \int h d(F - G) \right| \leq 2 \|h\|_V d_{ku}(F, G)$$

The asymptotics are covered by the following theorem. Here and in future Q_n denotes the empirical distribution associated with n i.i.d. random variables $(X_1(Q), \dots, X_n(Q))$ with common distribution Q .

Theorem 3.2 Let T be a real-valued functional which is locally uniformly linearizable on W . Then for all P in $W(\eta)$ there exist constants $\delta_1(P)$, $c_1(P)$ and $c_2(P)$ such that

$$\sup \left| \mathbf{P} \left(\sqrt{n} \left(\frac{T(Q_n) - T(Q)}{\sigma(T, Q)} \right) \leq u \right) - \Phi(u) \right| \leq c_1(P)(n^{-1/2} + r_n) \quad (7)$$

where

$$\sigma^2(T, Q) = \int I(u, T, Q)^2 dQ(u),$$

Φ is the distribution function of the standard $N(0, 1)$ distribution, r_n satisfies

$$\tau(\sqrt{-8 \log(r_n)/n}) = c_2(P)r_n/\sqrt{n}$$

and the supremum in (7) is taken over all $u \in \mathbb{R}$ and all $Q \in B(P, \delta_1(P))$.

PROOF. The proof is a standard application of the exponential inequality of Dvoretzky, Kiefer and Wolfowitz [14] and Berry-Esseen. It follows the proofs in Serfling [22] and Shao [23], Theorem 2.2 (ii), but with explicit consideration of the constants involved to ensure that they depend only on P for all Q and Q' in a sufficiently small neighbourhood of P . We omit the details. \square

We note that if $\tau(u) = u^2$ then $r_n = O(\log(n)/\sqrt{n})$ following corollaries.

Corollary 1 If $\tau(u) = u^2$ then

$$\sup_{u \in \mathbb{R}, Q \in B(P, \delta_2(P))} \left| \mathbf{P} \left(\sqrt{n} \left(\frac{T(Q_n) - T(Q)}{\sigma(T, Q)} \right) \leq u \right) - \Phi(u) \right| \leq c_3(P) \log(n)/\sqrt{n}.$$

Corollary 2 Suppose $\tau(u) = u^2$ and that the location-scale functional $T = (T_L, T_S)^t$ is locally uniformly linearizable on $W(\eta)$. Then for each $P \in W(\eta)$

$$\sup_{u \in \mathbb{R}, Q \in B(P, \delta_3(P))} \left| \mathbf{P} \left(n^{-1/2} (T_L(Q_n)) - T_L(Q) \leq u \sigma(T_L, Q_n) \right) - \Phi(u) \right| \leq c_4(P) \log(n) / \sqrt{n}.$$

We note that the error $O(\log n / \sqrt{n})$ is the same as that of Shao [23] Theorem 2.2 and that Corollary 2 justifies the approximations (3) and (4) for the approximation interval (2).

4 Locally uniformly linearizable M-functionals

4.1 M-functionals

The functionals we construct are weighted means with an automatic down-weighting of outlying observations. In order to accomplish this successfully we require a reliable method of determining which observations are outlying. Furthermore this must be done in a smooth manner so that the resulting weights are locally uniformly linearizable. The obvious solution is to use M-functionals.

An M-functional $M = (M_L, M_S)^t$ is the solution to the equations

$$\int \psi \left(\frac{u - M_L(P)}{M_S(P)} \right) dP(u) = 0 \quad (8)$$

$$\int \chi \left(\frac{u - M_L(P)}{M_S(P)} \right) dP(u) = 0. \quad (9)$$

We say that M is well-defined if (8) and (9) have a unique solution with $M_S(P) > 0$. To guarantee existence, uniqueness and locally uniform linearity we impose the following conditions on ψ and χ .

- M1 ψ is asymmetric, strictly increasing with $\psi(\infty) = 1$.
- M2 ψ has a continuous second derivative and $(1 + u^2)(|\psi^{(1)}(u)| + |\psi^{(2)}(u)|)$ is bounded and integrable
- M3 χ is symmetric, strictly increasing on $[0, \infty)$ with $\chi(0) = -1$ and $\chi(\infty) = 1$
- M4 χ has a continuous second derivative and $(1 + u^2)(|\chi^{(1)}(u)| + |\chi^{(2)}(u)|)$ is bounded and integrable
- M5 $\chi^{(1)}/\psi^{(1)}$ is strictly increasing.

Theorem 4.1 (Scholz [21]) *Let ψ and χ satisfy M1-M5 and P be a distribution with $\Delta(P) < 1/2$. Then the M -functional M is well-defined at P .*

It follows from Theorem 4.1 that M is well-defined on $W(1/2)$. A closer examination of the proof shows that M is well-defined if

$$p_1 + \Delta(P)\chi\left(\psi^{-1}\left(\frac{p_1 - p_2}{\Delta(P)}\right)\right) + p_2 > 0 \quad (10)$$

where p_1 and p_2 denote respectively the probabilities to the left and right of the largest atom.

The breakdown point $\epsilon^*(M, P)$ of M at P in $W(1/2)$ is defined by (5) where we set $M_L(Q) = \infty$ if M is not well-defined at Q . For ψ and χ satisfying M1-M5 we define ϵ_0 by

$$\psi^{-1}\left(\frac{\epsilon_0}{1 - \epsilon_0}\right) = \chi^{-1}\left(-\frac{\epsilon_0}{1 - \epsilon_0}\right) \quad (11)$$

Theorem 4.2 *For all P in $W(1/2)$ we have*

$$\min\{1/2 - \Delta(P), \epsilon_0\} \leq \epsilon^*(M, P) \leq (1 - \Delta(P))/2.$$

PROOF. If $\epsilon < 1/2 - \Delta(P)$ then for all $Q \in B(P, \epsilon)$ we have $\Delta(Q) < \Delta(P) + d_{ku}(Q, P) < 1/2$ so that $M(Q)$ is well-defined. Huber [18] pages 142-143 considers the breakdown of joint location and scale functionals for the gross error model. The method of proof for the lower bound continues to hold for the Kuiper neighbourhood. The condition $\epsilon < 1/2 - \Delta(P)$ guarantees that the scale does not implode and we obtain the left hand inequality of the theorem. The right hand side follows from general results for affinely equivariant functionals (see Davies [11]). \square

We note that it follows from the work of Clarke [9] that M is continuous at each point of $W(1/2)$. For any distribution $P \in W(1/2)$ we define the Jacobian $J(P)$ by

$$J(P) = -\frac{1}{s} \begin{pmatrix} \int \psi^{(1)}(u) dP'(u) & \int u\psi^{(1)}(u) dP'(u) \\ \int \chi^{(1)}(u) dP'(u) & \int u\chi^{(1)}(u) dP'(u) \end{pmatrix}$$

where $M(P) = (t, s)^t$ and $P'(B) = P(sB + t)$ for each Borel set B .

Lemma 4.1. *For each $P \in W(1/2)$ and for each $\epsilon(P) < \epsilon^*(M, P)$ we have*

$$0 < \inf \lambda_1(Q) \leq \sup \lambda_2(Q) < \infty$$

where $\lambda_1(Q)$ and $\lambda_2(Q)$ denote respectively the smallest and largest eigenvalues of the Jacobian $J(Q)$ and the infimum is taken over all $Q \in B(P, \epsilon(P))$.

PROOF. As $\epsilon < \epsilon^*$ it follows that $\inf M_S(Q) > 0$ implying that the elements of $J(Q)$ are bounded. This shows $\sup \lambda_2(Q) < \infty$. In order to show that the smallest eigenvalue is bounded away from zero we suppose this is not the case and consider a sequence $(Q_n)_{n=1}^\infty$ with $\lim_n |J(Q_n)| = 0$ where $|J|$ denotes the determinant of J . By Helly's Theorem there exists a convergence subsequence which tends to a possibly defective but non-degenerate distribution Q . If we denote this subsequence also by $(Q_n)_{n=1}^\infty$ then it follows from M2 and M4 that point masses at infinity have no effect on any of the integrals in the definition of $J(P)$. This implies $|J(Q)| = \lim_{n \rightarrow \infty} |J(Q_n)| = 0$ but as $Q(\mathbb{R}) > 1 - \epsilon(P) > 0$ the proof of Theorem 4.1. as given in Huber [18] shows that this cannot be the case. This contradiction proves the lemma. \square

The next step is to show that M is locally uniformly Lipschitz on $W(1/2)$.

Lemma 4.2. *For any distribution $P \in W(1/2)$ there exist constants $\delta_4(P) > 0$ and $c_5(P) > 0$ such that*

$$\|M(Q') - M(Q)\| \leq c_5(P) d_{ku}(Q', Q)$$

for all Q and Q' in $B(P, \delta_4(P))$.

PROOF: We set $\delta_5(P) = \min\{1/2 - \Delta(P), \epsilon_0\}/2$ with ϵ_0 as given by (11). This implies

$$c_6(P) = \sup\{M_S(Q)^{-1} : Q \in B(P, \delta_5(P))\} < \infty. \quad (12)$$

For Q and Q' in $B(P, \delta_5(P))$ we write

$$M(Q) = (t, s)^t : M(Q') = (t', s')^t.$$

On writing

$$y = \frac{u-t}{s} : y' = \frac{u-t'}{s'} : \eta = \frac{t-t'}{s} : \kappa = \frac{s-s'}{s}$$

a Taylor expansion gives

$$\psi(y') - \psi(y) = \eta \psi^{(1)}(y) + \kappa y \psi^{(1)}(y) + O_P(\|M(Q') - M(Q)\|^2). \quad (13)$$

On integrating (13) with respect to Q' and using Theorem 3.1 and M2 we obtain

$$\begin{aligned} & \int \psi(y) dQ'(u) \\ &= \eta \int \psi^{(1)}(y) dQ(u) + \kappa \int y \psi^{(1)}(y) dQ(u) \\ & \quad O_P(\|M(Q') - M(Q)\| (d_{ku}(Q', Q) + \|M(Q') - M(Q)\|)) \end{aligned}$$

with the same inequality with χ in place of ψ .

This gives with $\Psi = (\psi, \chi)^t$

$$\begin{aligned} \int \Psi(y)dQ'(u) &= J(Q) \left(M(Q') - M(Q) \right) + \\ &\quad O_P(\|M(Q') - M(Q)\|(\|M(Q') - M(Q)\| + d_{ku}(Q', Q))) \end{aligned}$$

for all Q and Q' in $B(P, \delta_6(P))$. A further application of Theorem 3.1 shows that

$$\left\| \int \Psi(y)dQ'(u) \right\| = \left\| \int \Psi(y)d(Q'(u) - Q(u)) \right\| \leq c_7(P)d_{ku}(Q', Q)$$

and hence

$$J(Q) \left(M(Q') - M(Q) \right) = R_8 \tag{14}$$

with

$$|R_8| \leq c_8(P) \left(\|M(Q') - M(Q)\|^2 + d_{ku}(Q', Q) \right).$$

From this and Lemma 4.1 we can conclude that

$$\|M(Q') - M(Q)\| \leq \Lambda(P)^{-1}c_9(P) \left(\|M(Q') - M(Q)\|^2 + d_{ku}(Q', Q) \right)$$

Because of the continuity of M we can choose $0 < \delta_7(P) \leq \delta_6(P)$ such that

$$\|M(Q') - M(Q)\| \leq \frac{1}{2}\Lambda(P)c_9(P)^{-1}$$

for all Q and Q' in $B(P, \delta_8(P))$ it follows that

$$\|M(Q') - M(Q)\| \leq 2\Lambda(P)^{-1}c_9(P)d_{ku}(Q', Q)$$

as was to be proved. \square

Theorem 4.3 *Suppose that M1-M5 are satisfied. Then the functional M is locally uniformly linearizable on $W(1/2)$ with $\tau(u) = u^2$. In particular for each $P \in W(1/2)$ there exist constants $\delta_9(P)$ and $c_{10}(P)$ such that*

$$\left\| M(Q') - M(Q) - J(Q)^{-1} \int \Psi \left(\frac{u - M_L(Q)}{M_S(Q)} \right) dQ'(u) \right\| \leq c_{10}(P)d_{ku}(Q', Q)^2 \tag{15}$$

for all Q and Q' in $B(P, \delta_{10}(P))$ and where $\Psi = (\psi, \chi)^t$.

PROOF: The inequality (15) follows from what we have just proved. The locally uniform linearity follows from (15) by showing that the function

$$I(u, M, Q) = J(Q)^{-1} \Psi \left(\frac{u - M_L(Q)}{M_S(Q)} \right)$$

satisfies L1 and L2. This is done using arguments similar to those of the proof of Lemma 4.2. \square

5 A smooth weighted average

5.1 Locally uniform linearity

As mentioned in the introduction the functional we construct is a weighted mean with outlying observations being downweighted. The outlyingness of observations is determined using the M-functional of the last section. In order to retain the locally uniform linearity we require a smooth weight function w . We shall assume that it fulfills the following conditions:

W1 w is symmetric and strictly decreasing on $[0, \infty)$

W2 $w(1) = 1, w(\infty) = 0$

W3 w has a continuous second derivative $w^{(2)}$

W4 $\lim_{u \rightarrow \infty} u^2 (w(u) + |w^{(1)}(u)| + |w^{(2)}(u)|) = 0$

Given the weight function w we define the weighted mean functional $M^w = (M_L^w, M_S^w)^t$ by

$$M_L^w(P) = \frac{\int u w \left(\frac{u - M_L(P)}{M_S(P)} \right) dP(u)}{\int w \left(\frac{u - M_L(P)}{M_S(P)} \right) dP(u)} \quad (16)$$

and

$$M_S^w(P)^2 = \frac{\int (u - M_L^w(P))^2 w \left(\frac{u - M_L(P)}{M_S(P)} \right) dP(u)}{\int w \left(\frac{u - M_L(P)}{M_S(P)} \right) dP(u)} \quad (17)$$

It is clear that M^w is well-defined on $W(1/2)$ and that it is affinely equivariant. We show firstly that its breakdown point is at least that of M .

Theorem 5.1 *For any P in $W(1/2)$ we have*

$$\epsilon^*(M^w, P) \geq \epsilon^*(M, P).$$

PROOF: Choose an $\epsilon < \epsilon^*(M, P)$ and an interval I with

$$\sup_{u \in I, Q \in B(P, \epsilon)} \left| \frac{u - M_L(Q)}{M_S(Q)} \right| = A < \infty \quad (18)$$

and hence for all $Q \in B(P, \epsilon)$ we have

$$\int w \left(\frac{u - M_L(Q)}{M_S(Q)} \right) dQ(u) \geq w(A) Q(I) \geq w(A) \left(\frac{3}{4} - \epsilon \right) > 0. \quad (19)$$

To show that $M_L^w(Q)$ remains bounded we note that

$$\sup_{Q \in B(P, \epsilon)} \left| \int u w \left(\frac{u - M_L(Q)}{M_S(Q)} \right) dQ(u) \right| < \infty$$

which follows from

$$\int uw \left(\frac{u - M_L(Q)}{M_S(Q)} \right) dQ(u) = M_L(Q) \int w \left(\frac{u - M_L(Q)}{M_S(Q)} \right) dQ(u) + M_S(Q) \int \left(\frac{u - M_L(Q)}{M_S(Q)} \right) w \left(\frac{u - M_L(Q)}{M_S(Q)} \right) dQ(u),$$

the boundedness of $M_L(Q)$ and $M_S(Q)$ and W4.

A similar argument shows that $M_S^w(Q)$ is bounded above and it remains to show that it is bounded away from zero. If this is not the case then there exists a sequence $(Q_n)_1^\infty$ in $B(P, \epsilon)$ with

$$\lim_{n \rightarrow \infty} \int (u - M_L^w(Q_n))^2 w \left(\frac{u - M_L(Q_n)}{M_S(Q_n)} \right) dQ_n(u) = 0.$$

By choosing subsequences if necessary we can assume that

$$\lim_{n \rightarrow \infty} (M_L^w(Q_n), M_L(Q_n), M_S(Q_n)) = (t', t, s)$$

with $s > 0$. Let $\gamma > 0$ and $\eta > 0$ be given. We set $I'(\gamma) = [t' - \gamma, t' + \gamma]$ and choose a finite interval $I(\eta)$ with $P(I(\eta)) > 1 - \eta$ and $I(\eta) \supset I'(\gamma)$. It follows that $\lim_{n \rightarrow \infty} Q_n(I(\eta) \cap (\mathbb{R} \setminus I'(\gamma))) = 0$ and hence

$$\lim_{n \rightarrow \infty} (Q_n(I(\eta)) - Q_n(I'(\gamma))) = 0.$$

As $Q_n(I(\eta)) > P(I(\eta)) - \epsilon > 1 - \eta - \epsilon$ we obtain

$$\limsup_{n \rightarrow \infty} Q_n(I'(\gamma)) > 1 - \eta - \epsilon.$$

This implies

$$P(I'(\gamma)) \geq \limsup_{n \rightarrow \infty} Q_n(I'(\gamma)) - \epsilon > 1 - \eta - 2\epsilon.$$

On letting γ tend to zero it follows that $\Delta(P) \geq P(\{t'\}) \geq 1 - \eta - 2\epsilon$ and as η is arbitrary we obtain $\Delta(P) \geq 1 - 2\epsilon > \Delta(P)$ by Theorem 4.2. This is a contradiction and proves the theorem. \square

Theorem 5.2 *Suppose M1-M5 and W1-W4 are satisfied. Then the functional M^w is locally uniformly linearizable with $\tau(u) = u^2$.*

PROOF: This follows from a second order Taylor expansion using W1-W4 and the locally uniform linearity of M . \square

5.2 The influence function

If the approximate approximation region is to be calculated using asymptotics then the influence function $I(\cdot, M^w, P)$ is required. We give it here in terms of the influence function $I(\cdot, M, P)$ of the M-functional. To shorten the expressions we set

$$u' = \frac{u - M_L(P)}{M_S(P)}, \quad u'' = \frac{u - M_L^w(P)}{M_S(P)}, \quad u''' = \frac{u - M_L^w(P)}{M_S^w(P)}.$$

We write

$$\begin{aligned} w(P) &= \int w(u') dP(u) \\ C_1(P) &= \frac{1}{w(P)} \int u'' w^{(1)}(u') dP(u) \\ C_2(P) &= \frac{1}{w(P)} \int u'' u' w^{(1)}(u') dP(u) \\ C_3(P) &= -\frac{1}{w(P)} \int u''' w(u') dP(u) \\ C_4(P) &= \frac{1}{2w(P)} \int (1 - u''^2) w^{(1)}(u') dP(u) \\ C_5(P) &= \frac{1}{2w(P)} \int (1 - u''^2) u' w^{(1)}(u') dP(u) \end{aligned}$$

With this notation we have

$$\begin{aligned} I(u, M_L^w, P) &= \frac{(u - M_L^w(P))}{w(P)} w\left(\frac{u - M_L(P)}{M_S(P)}\right) \\ &\quad - C_1(P) I(u, M_L, P) - C_2(P) I(u, M_S, P) \end{aligned}$$

and

$$\begin{aligned} I(u, M_S^w, P) &= \frac{M_S^w(P)}{2w(P)} \left(\left(\frac{u - M_L^w(P)}{M_S^w(P)} \right)^2 - \right) w\left(\frac{u - M_L(P)}{M_S(P)}\right) + \\ &\quad C_3(P) I(u, M_L^w, P) + C_4(P) I(u, M_L, P) + C_5(P) I(u, M_S, P) \end{aligned}$$

6 A second construction:coarsening

6.1 Coarsening

We write ρ for a function which satisfies the following conditions:

R1 $\rho : [-1, 1] \rightarrow \mathbb{R}$ is symmetric, positive and non-increasing on $[0, 1)$ with $\rho(0) = 1$ and $\rho(1) = 0$.

R2 ρ has a continuous second derivative with $\rho^{(1)}(1) = 0$ and $\rho^{(2)}(1) = 0$.

We recall the definition (6) of \tilde{F} for the distribution function F . For any distribution P we denote the associated distribution function by F_P . With this notation and for any integer $m \geq 3$ we define

$$y(j, m, P) = \frac{\int u \rho \left(2(m+1) \left(\tilde{F}_P(u) - \frac{j}{m+1} \right) \right) dF_P(u)}{\int \rho \left(2(m+1) \left(\tilde{F}_P(u) - \frac{j}{m+1} \right) \right) dF_P(u)}. \quad (20)$$

Clearly $y(j, m, P)$ is well-defined if

$$(21)$$

and

$$(22)$$

Lemma 6.1. $y(j, m, P)$ is well-defined for all $P \in W(\frac{1}{m+1})$.

PROOF: As the integral in (22) is over a finite interval it is finite. To prove (21) we note that if it does not hold for some j then for all u either $\tilde{F}_P(u) \geq \frac{j+\frac{1}{2}}{m+1}$ or $\tilde{F}_P(u) \leq \frac{j-\frac{1}{2}}{m+1}$. This implies

$$\Delta(P) \geq \frac{j + \frac{1}{2}}{m+1} - \frac{j - \frac{1}{2}}{m+1} = \frac{1}{m+1}$$

contradicting $P \in W(\frac{1}{m+1})$. \square

Theorem 6.1 The functional $y(j, m, \cdot) : W(\frac{1}{m+1}) \rightarrow \mathbb{R}$ is locally uniformly linearizable.

PROOF: We set $s = 2(m+1)$, $t = \frac{j}{m+1}$, $y(P) = y(j, m, P)$, $y(Q) = y(j, m, Q)$ and write F and G for the distribution functions of P and Q respectively. We have

$$y(Q) - y(P) = \frac{\int (u - y(P)) \rho \left(s(\tilde{G}(u) - t) \right) dG(u)}{\int \rho \left(s(\tilde{G}(u) - t) \right) dG(u)}. \quad (23)$$

It suffices to show that the numerator and denominator of (20) are locally uniformly linearizable. We consider only the numerator. Precise arguments can be given just as in Section 4 but we restrict ourselves to the main steps. A Taylor expansion gives

$$\int (u - y(P)) \rho \left(s(\tilde{G}(u) - t) \right) dG(u) = R_1 + R_2 + R_3$$

where

$$\begin{aligned}
R_1 &= \int (u - y(P))\rho\left(s(\tilde{F}(u) - t)\right) d(G - F)(u) \\
R_2 &= s \int (u - y(P))\rho^{(1)}\left(s(\tilde{F}(u) - t)\right) \left(\tilde{G}(u) - \tilde{F}(u)\right) dG(u) \\
R_3 &= \frac{s^2}{2} \int (u - y(P))\rho^{(2)}\left(s(\tilde{F}(u, \theta) - t)\right) \left(\tilde{G}(u) - \tilde{F}(u)\right)^2 dG(u).
\end{aligned}$$

with $\tilde{F}(u, \theta) = \theta(u)\tilde{F}(u) + (1 - \theta(u))\tilde{G}(u)$ for some $\theta(u), 0 < \theta(u) < 1$. Clearly $R_3 = O_P(d_{ku}(Q, P)^2)$ whilst for R_2 we have

$$\begin{aligned}
R_2 &= s \int (u - y(P))\rho^{(1)}\left(s(\tilde{F}(u) - t)\right) \left(\tilde{G}(u) - \tilde{F}(u)\right) dF(u) + \\
&\quad \int (u - y(P))\rho^{(1)}\left(s(\tilde{F}(u) - t)\right) \left(\tilde{G}(u) - \tilde{F}(u)\right) d(G - F)(u) \\
&= s \int (u - y(P))\rho^{(1)}\left(s(\tilde{F}(u) - t)\right) \left(\tilde{G}(u) - \tilde{F}(u)\right) dF(u) + \\
&\quad O_P(d_{ku}(Q, P)^2)
\end{aligned}$$

On noting

$$\begin{aligned}
&\int (u - y(P))\rho^{(1)}\left(s(\tilde{F}(u) - t)\right) \left(\tilde{G}(u) - \tilde{F}(u)\right) dF(u) \\
&= \int \left(\int (u - y(P))\rho^{(1)}\left(s(\tilde{F}(u) - t)\right) (\{v \leq u\} - \frac{1}{2}\{u = v\}) dF(u) \right) (G - F)(v)
\end{aligned}$$

and on performing a similar calculation for the denominator of (23) we finally obtain

$$y(Q) - y(P) = \int I(u, j, m, P)d(Q - P)(u) + O_P(d_{ku}(Q, P)^2)$$

where

$$I(u, j, m, P) = \frac{I'(u, j, m, P) - \int I'(u, j, m, P)dF(u)}{\int \rho\left(s(\tilde{F}(u) - t)\right) dF(u)}.$$

and

$$\begin{aligned}
I'(u, j, m, P) &= (u - y(P))\rho\left(s(\tilde{F}(u) - t)\right) - \\
&\quad \frac{s}{2}(u - y(P))\rho^{(1)}\left(s(\tilde{F}(u) - t)\right) - \\
&\quad s \int (v - y(P))\rho^{(1)}\left(s(\tilde{F}(v) - t)\right) \{v \geq u\} dF(v)
\end{aligned}$$

It can be checked that the functions $I(\cdot, j, m, \cdot)$ fulfill the conditions L1 and L2 proving the theorem. \square

6.2 The location and scale functionals

We assume m to be an odd number and write $m_e = \frac{(m+1)}{2}$. The location functional T_L^c is simply defined by

$$T_L^c(P) = y(m_e, m, P)$$

From what was shown in the last section it is clear that T_L^c is locally uniformly linearizable with influence function $I(u, T_L^c, P) = I(u, m_e, m, P)$. To obtain its breakdown behaviour we note that the definition of $y(m_e, m, P)$ depends only on an interval centred at the median with a mass of at most $\frac{1}{(m+1)}$. This implies

Theorem 6.2 *For any $P \in W(\frac{1}{m+1})$ and for any $\epsilon < \frac{1}{2} - \frac{1}{m+1}$ we have*

$$\sup_Q |T_L^c(Q)| < \infty$$

where the supremum is taken over all $Q \in B(P, \epsilon) \cap W(\frac{1}{m+1})$.

Given the location functional T_L^c we define the scale functional T_S^c to be the unique solution of

$$\sum_{j=1}^m \chi \left(\frac{y(j, m, P) - y(m_e, m, P)}{T_S^c(P)} \right) = 0$$

where the function χ satisfies M3 and M4. The locally uniform linearity of T_S^c is covered by

Theorem 6.3 *The scale functional T_S^c is locally uniformly linearizable on $W(\frac{1}{m+1})$ and satisfies*

$$\frac{T_S^c(Q) - T_S^c(P)}{T_S^c(P)} = \frac{\int I_S(u, P)(Q - P)(u)}{H} + O_P(d_{ku}(Q, P)^2)$$

where

$$I_S(u, P) = \sum_{j=1}^m \chi^{(1)} \left(\frac{y(j, m, P) - T_L^c(P)}{T_S^c(P)} \right) \left(\frac{I(u, j, m, P) - I(u, m_e, m, P)}{T_S^c(P)} \right)$$

and

$$H = \sum_{j=1}^m \chi^{(1)} \left(\frac{y(j, m, P) - T_L^c(P)}{T_S^c(P)} \right) \left(\frac{y(j, m, P) - y(m_e, m, P)}{T_S^c(P)} \right)$$

PROOF: The proof is just a straightforward calculation. \square .

The breakdown behaviour of T_S^c is covered by the following theorem.

Theorem 6.4 (a) For any $P \in W(\frac{1}{m+1})$ and for any $\epsilon < \frac{1}{2} - \frac{1}{m+1}$ we have

$$\sup_Q T_S^c(Q) < \infty$$

where the supremum is taken over all $Q \in B(P, \epsilon) \cap W(\frac{1}{m+1})$.

(b) For any $P \in W(\frac{1}{m+1})$ and for any $\epsilon < \frac{1}{2} - \frac{2}{m+1} - \Delta(P)$ we have

$$\inf_Q T_S^c(Q) > 0$$

where the infimum

PROOF: (a) Let ϵ be as in (a) of the theorem and $(Q_n)_{n=1}^\infty$ be a sequence of distributions in $B(P, \epsilon) \cap W(\frac{1}{m+1})$ with distribution functions $(G_n)_{n=1}^\infty$. We suppose that $y(i, m, Q_n)$ tends to $-\infty$ and that $y(j, m, Q_n)$ tends to ∞ . We set $a_n = G_n^{-1}(\frac{i}{m+1} - \frac{1}{2(m+1)})$ and $b_n = G_n^{-1}(\frac{j}{m+1} + \frac{1}{2(m+1)})$. Then the a_n tend to $-\infty$ and the b_n to ∞ . As $\lim P((a_n, b_n)) = 1$ we conclude

$$\begin{aligned} \epsilon &> 1 - (G_n(b_n) - G_n(a_n)) \\ &\geq 1 - \left(\frac{j}{m+1} + \frac{1}{2(m+1)} - \left(\frac{i}{m+1} - \frac{1}{2(m+1)} \right) \right) \\ &= \frac{m+1-j+i-1}{m+1} \end{aligned}$$

It follows that $m+1-j+i < \frac{m+1}{2}$ and hence the total number of the y 's which tend to $\pm\infty$ is at most $\frac{m+1}{2} - 1$ which is strictly less than $\frac{m}{2}$. This implies that the $T_S^c(Q_n)$ remain bounded proving (a) of the theorem.

To prove (b) we set $a_n = G_n^{-1}(\frac{i}{m+1} + \frac{1}{2(m+1)})$ and $b_n = G_n^{-1}(\frac{j}{m+1} - \frac{1}{2(m+1)})$ and suppose that the $T_L^c(Q_n) - y(i, m, Q_n)$ and $y(j, m, Q_n) - T_L^c(Q_n)$ converge to some finite number ζ . An argument similar to the one just used shows that

$$\begin{aligned} \epsilon &> |P(\{\zeta\}) - (G_n(b_n) - G_n(a_n))| \\ &\geq \left(\frac{j}{m+1} - \frac{1}{2(m+1)} - \left(\frac{i}{m+1} + \frac{1}{2(m+1)} \right) \right) - \Delta(P) \\ &= \frac{j-i-1}{m+1} - \Delta(P) \end{aligned}$$

It follows that $j-i+1 < \frac{m+1}{2}$ and hence the total number of the $(y - T_L^c(Q_n))$'s which tend to zero is at most $\frac{m+1}{2} - 1$ which is strictly less than $\frac{m}{2}$. This implies that the scale cannot implode. \square .

6.3 A smooth weighted average

Just as in Section 5 we take the final functionals to be weighted averages with the weights being determined by the functionals T_L^c and T_S^c . The Theorems 5.1 and 5.2 hold correspondingly as do the expressions for the influence functions given in Section 5.2.

6.4 Adjusting the weights

As indicated above one further possible use of the coarsened sample is to allow adjustment of the weights depending on the shape of the sample. We do this by defining the standardized coarsened sample by

$$y^s(j, m, P) = \frac{y(j, m, P) - y(m_e, m, P)}{T_S^c(P)}$$

The standardized coarsened sample is a locally uniformly differentiable approximation to the distribution from which it is calculated. The nature of the approximation may be seen by the following examples for the Gauss and slash distributions. The and U uniformly distributed on $[0, 1]$ and independent of X . Morgenthaler and Tukey [19] discuss its advantages over the Cauchy distribution for modelling heavy tailed distributions. We take $m = 15$ and for reasons of symmetry we give only the positive values.

Gauss:	0.157	0.319	0.489	0.674	0.887	1.151	1.537
slash:	0.136	0.280	0.441	0.636	0.906	1.387	2.843

For general m let $y^s(j, m, G)$, $1 \leq j \leq m$, and $y^s(j, m, C)$, $1 \leq j \leq m$, denote the coarsened samples Gauss and slash distributions respectively. For any sample $\mathbf{x}_n = (x_1, \dots, x_n)$ let $y^s(j, m, \mathbf{x}_n)$, $1 \leq j \leq m$, denote the coarsened sample. Distances between the empirical coarsened sample and those derived from the Gauss and slash distributions may be defined as follows;

$$d(\mathbf{x}_n, G)^2 = \sum_{j=1}^m (y^s(j, m, \mathbf{x}_n) - y^s(j, m, G))^2 \quad (24)$$

$$d(\mathbf{x}_n, S)^2 = \sum_{j=1}^m (y^s(j, m, \mathbf{x}_n) - y^s(j, m, S))^2. \quad (25)$$

$$(26)$$

Other choices are possible. If now a weight function w_G is judged to be appropriate for Gaussian samples and the weight function w_S for slash samples we may now use a convex combination based on (24) and take the weight function to be

$$w = \frac{d(\mathbf{x}_n, S)w_G + d(\mathbf{x}_n, G)w_S}{d(\mathbf{x}_n, S) + d(\mathbf{x}_n, G)}$$

7 Examples

7.1 Choice of functions

We restrict attention to the weighted mean with the weights being determined by the M-functional. To define the M-functional we must specify defining functions ψ and χ which satisfy the conditions imposed in Section 3. The functions we use are of the form

$$\psi(u) = \psi(u, cl) = \frac{\exp(u/cl) - 1}{\exp(u/cl) + 1}$$

and

$$\chi(u) = \frac{u^2 - 1}{u^2 + 1}.$$

It is clear that M1-M4 are satisfied. Numerical calculations show that M5 is satisfied if $cl < 0.6418$

Apart from a possible uncontrollable atom of P the breakdown point is effectively governed by the ϵ_0 of (11). We choose $cl = 0.2$ giving $\epsilon_0 = 0.427$. The weight function w is taken to be

$$w(u) = \exp\left(-\left(\frac{u}{5.0}\right)^2\right)$$

This satisfies W1-W4. The resulting functional is now completely specified and will be denoted by $M^w = (M_L^w, M_S^w)$.

7.2 Comparisons

As argued in Section 2 a procedure should be compared with other competing procedures where the comparison is to be as broadly based as possible using theoretical challenges and real data sets. A comparison on the scale required is beyond the bounds of this article but to indicate what is intended we give the results of a small study. We consider three alternative procedures. They are (i) the mean and standard deviation, (ii) the median and the MAD and (iii) the mean and standard deviation after elimination of outliers as proposed by Hampel [15]. This involves eliminating all observations which deviate from the median by more than $5.2\text{MAD}(\mathbf{x}_n)$ where $\text{MAD}(\mathbf{x}_n)$ denotes the median absolute deviation of the data \mathbf{x}_n . We shall denote these proposals by mean/sdv, median/MAD and Hmean/Hsdv respectively. None of the proposals is locally uniformly linearizable but this is least likely to be felt for Hmean/Hsdv.

7.3 Theoretical challenges

In this section we consider three theoretical challenges in the sense of Morgenthaler and Tukey [19]. It is not at all clear which theoretical challenges we should

Table 1: Gauss

n	20	50	100	∞
M_L^w	2.08	1.99	2.02	2.03
mean	1.96	1.96	1.96	1.96
median	2.38	2.41	2.47	2.46
Hmean	2.02	1.99	1.96	1.96

Table 2: Slash

n	20	50	100	∞
M_L^w	4.96	4.86	4.76	4.68
mean	72.33	131.77	157.00	∞
median	4.96	4.86	4.95	4.91
Hmean	5.79	5.54	5.61	5.36

use. Traditionally the Gaussian challenge is included as is one with heavy tails which we take to be the slash distribution rather than the Cauchy distribution for reasons given in Morgenthaler and Tukey [19]. Many real data sets exhibit a high degree of discreteness even though they would normally be modelled using a continuous distribution. For this reason we include the die challenge by which we mean the uniform distribution over the integers $\{1, \dots, 6\}$.

We restrict consideration to the location part of the functionals without any studentization. This enables us to compare the location parts directly without the influence of the scale functionals. The comparisons will be made using the 0.975-quantile of $\sqrt{n}(T_L(P_n(P)) - T_L(P))$ for $n = 20, 50, 100, \infty$. The results for the three challenges are given in tables 1, 2 and 3. The mean fails for the slash challenge and the median for the die.

Table 3: Die

n	20	50	100	∞
M_L^w	4.10	3.85	3.86	3.97
mean	3.35	3.28	3.33	3.35
median	6.71	3.54	5.00	∞
Hmean	3.35	3.28	3.33	3.35

Table 4: Moth

0	124	5	4	10	2	15	1	22	1
1	45	6	5	11	6	16	2	23	1
2	28	7	8	12	2	18	1	30	1
3	18	8	3	13	1	20	1	31	2
4	14	9	3	14	1	21	1	60	1

Table 5: Study

8,0	1	12,1	26	14,3	4	17,1	3	20,2	1
8,1	5	12,2	7	15,0	1	17,2	1	20,4	1
9,1	10	12,5	1	15,1	8	18,0	1	21,0	1
9,3	1	13,0	3	15,3	2	18,1	1	22,0	2
10,1	18	13,1	14	15,3	5	18,3	1	22,3	2
10,2	3	13,2	7	16,1	6	18,5	1	22,5	2
10,5	1	13,4	1	16,2	1	19,1	1	23,1	2
11,1	14	13,5	1	16,3	2	19,3	2		
11,2	3	14,1	10	16,5	1	19,5	2		
11,4	1	14,2	3	17,0	1	20,1	4		

7.4 Practical challenges

The three practical challenges we consider are the moth data, the study data and Darwin's data on self- and cross-fertilization. The complete data sets are given below. The moth data give the number of moths caught on each of 276 nights, the first number in each column is the number of moths whilst the second gives the frequency.

The second example gives the lengths of study of 189 German students. The first number gives the lengths of study to the nearest month or one-sixth of one semester. Thus the number 10,3 is to be read as $10\frac{3}{6}$ semesters ie as 10.5 semesters.

Finally we include Darwin's data which gives the differences in heights of self and cross-fertilized plants. The unit of measurement is one-eighth of an inch

Table 6: Darwin

-67	-48	6	8	14	16	23	24
28	29	41	49	56	60	75	

Table 7: Moth: Approximation intervals (27)

Functional	M^w	mean/sdv	median/mad	Hmean/Hsdv
Location/scale	1.03/1.57	2.93/6.08	1.00/1.00	1.1/1.5
Asy.boot	[0.72,1.35]	[2.21,3.65]	[1.00,1.00]	[0.83,1.36]
boot	[0.75,1.43]	[2.34,3.89]	[1.00,∞]	[0.92,∞]
Gauss	[0.82,1.24]	[2.21,3.65]	[0.78,1.22]	[0.93,1.27]
Slash	[0.84,1.23]	[2.29,3.56]	[0.80,1.20]	[0.90,1.31]

(Andrews and Herzberg [1]). The data set has been analysed several times in the literature (Barnard [3], Box and Tiao [8], Sprott [24]).

The problem with practical challenges is that it is not possible to consider approximation intervals based on all models which are a reasonable approximation to the data as defined in Section 2. Any given challenge can of course be included. We restrict ourselves to the standard bootstrap obtained by simulations (3) and the asymptotic bootstrap (4). In the sense of Morgenthaler and Tukey we also include the Gauss and the slash challenges. Apart from the Darwin data these are not adequate approximations but it may be argued that if the approximation intervals based on these challenges are not too different from the bootstrap intervals then this is further evidence of stability.

The statistic we use is the studentized location functionals given by

$$M_{LS}^w(P_n(P), P) = (M_L^w(P_n(P)) - M_L^w(P)) / M_S^w(P_n(P))$$

with M_L^w and M_S^w given by (16) and (17) respectively and where the functions ψ , χ and w are as in Section 7.1. The measures of performance we employ are the size and variability of approximation intervals of the form

$$\begin{aligned} & [M_L^w(P_n(\mathbf{x}_n)) - qu(n, 0.975, M_{LS}^w, P)M_S^w(P_n(\mathbf{x}_n)) / \sqrt{n}, \\ & M_L^w(P_n(\mathbf{x}_n)) - qu(n, 0.025, M_{LS}^w, P)M_S^w(P_n(\mathbf{x}_n)) / \sqrt{n}] \end{aligned} \quad (27)$$

for various P . Here as before $qu(n, \alpha, M_{LS}^w, P)$ denotes the α quantile of the distribution of the studentized location functional $M_{LS}^w(P_n(\mathbf{x}_n), P)$. The models P we consider are $P = P_n(\mathbf{x}_n)$ (bootstrap), Gauss and slash. The asymptotic bootstrap approximation interval uses the quantile $qu(\infty, \alpha, M_{LS}^w, P)$.

7.5 Conclusions

Taking the theoretical challenges and the real data sets together we see that the weighted mean based on the M-functional is the only one which does not fail completely in the sense of giving infinite approximation intervals. Its worst performance is perhaps for the die distribution where the approximation intervals are 20% longer than those of the mean. The approximation intervals for

Table 8: Study lengths: Approximation intervals (27)

Functional	M^w	mean/sdv	median/mad	Hmean/Hsdv
Location/Scale	13.12/2.88	13.59/3.45	13.17/2.00	13.59/3.45
Asy.boot	[12.61,13.62]	[13.09,14.08]	[13.17,13.17]	[13.09,14.08]
boot	[12.65,13.65]	[13.12,14.12]	[13.01,14.17]	[13.11,14.49]
Gauss	[12.66,13.59]	[13.09,14.08]	[12.62,13.68]	[13.08,14.09]
Slash	[12.68,13.56]	[13.15,14.02]	[12.68,13.65]	[13.03,14.14]

Table 9: Darwin: Approximation intervals (27)

Functional	M^w	mean/sdv	median/mad	Hmean/Hsdv
Location/Scale	26.62/27.00	20.93/36.47	24.00/17.00	27.21/28.86
Asy.boot	[13.02,40.22]	[2.47,39.39]	[24.00,24.00]	[9.78,44.66]
boot	[9.52,43.05]	[-7.81,38.15]	[4.73,42.89]	[5.55,51.59]
Gauss	[8.44,45.29]	[0.73,41.13]	[4.68,45.13]	[8.13,46.29]
Slash	[9.40,43.23]	[2.91,38.95]	[7.17,40.83]	[7.61,46.81]

the real data sets show an acceptable degree of stability over the four different ways of calculating them. The mean/sdv is of course susceptible to outliers as for example in the moth data whereas the median/mad is susceptible to atoms as in the moth and study length data where the approximation intervals are not stable. The Hmean/Hsdv functional performs well apart from the bootstrap approximation interval for the moth data. Its simplicity is a great advantage and for this reason it may well be preferable to M^w if large atoms are not to be expected. It can certainly be recommended for a first course in data analysis as an alternative to the usual mean/sdv functional.

8 Acknowledgement

I acknowledge the help of an editor and two anonymous referees whose comments improved the contents, clarity and the presentation of the paper.

9 References

References

- [1] Andrews, D.F. and Herzberg, A.M. (1985) *Data: A collection of problems from many fields for the student and research worker*. Springer, New York.
- [2] Bahadur, R.R. and Savage, J.S. (1955) The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics* 27, 1115-1122.
- [3] Barnard, G.A. (1987) What influences influence? *J. Appl. Statist.* 14, 107-116.
- [4] Bednarski, T., Clarke, B.R. and Kolkiewicz, W. (1991) Statistical expansions and locally uniform Fréchet differentiability. *J. Austral. Math. Soc.* 50, 88-97.
- [5] Bickel, P.J. and Lehmann, E.L. (1975a) Descriptive statistics for nonparametric models I. Introduction. *Annals of Statistics* 3, 1038-1044.
- [6] Bickel, P.J. and Lehmann, E.L. (1975b) Descriptive statistics for nonparametric models II. Location. *Annals of Statistics* 3, 1045-1069.
- [7] Bickel, P.J. and Lehmann, E.L. (1976) Descriptive statistics for nonparametric models III. Dispersion. *Annals of Statistics* 4, 1139-1158.
- [8] Box, G.E.P. and Tiao, G.C. (1973) *Bayesian inference in statistical analysis*. Addison-Wesley, Reading, Mass.
- [9] Clarke, B.R. (1983) Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Annals of Statistics* 4, 1196-1205.
- [10] Croux, C. (1994) Efficient high-breakdown estimators of scale. *Stat. Prob. Letters* 19, 371-369.
- [11] Davies, P.L. (1993) Aspects of robust linear regression. *Annals of Statistics* 21, 1843-1899.
- [12] Davies, P.L. (1995) Data features. *Statistica Neerlandica*, 49, 185-245
- [13] Dudley, R.M. (1989) *Real Analysis and Probability*. Wadsworth and Brooks/Cole,
- [14] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956) Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals of Mathematical Statistics* 27, 642-669.

- [15] Hampel, F.R. (1985) The breakdown points of the mean combined with some rejection rules. *Technometrics* 27, 95-107.
- [16] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986) *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
- [17] Huber, P.J. (1964) Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73-101.
- [18] Huber, P.J. (1981) *Robust Statistics*. Wiley, New York.
- [19] Morgenthaler, S. and Tukey, J.W. (1991) *Configural Polysampling: A Route to Practical Robustness*. Wiley, New York.
- [20] Rieder, H. (1994) *Robust Asymptotic Statistics*. Springer, Berlin.
- [21] Scholz, F.W. (1971) Comparison of optimal location estimators. Ph.D. thesis, Department of Statistics, University of California, Berkeley.
- [22] Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- [23] Shao, J. (1989) Functional calculus and asymptotic theory for statistical analysis. *Statistics and Probability Letters* 8, 397-405.
- [24] Sprott, D.A. (1978) Robustness and non-parametric procedures are not the only or the safe alternatives to normality. *Canad. J. Psychol./Rev.Canad. Psychol.* 32, 180-185.