Higher order asymptotics for the MSE of M-estimators on shrinking neighborhoods

Peter Ruckdeschel*
Mathematisches Institut
Universität Bayreuth
D-95440 Bayreuth
Germany
e-Mail: peter.ruckdeschel@uni-bayreuth.de

3rd November 2005

Abstract

In the setup of shrinking neighborhoods about an ideal central model, Rieder (1994) determines the as. linear estimator minimaxing MSE on these neighborhoods. We address the question to which degree this as. optimality carries over to finite sample size. We consider estimation of a one-dim. location parameter by means of M-estimators S_n with monotone influence curve ψ. Using Donoho and Huber (1983)'s finite sample breakdown point ε_0 for S_n, we define thinned out convex contamination balls $\tilde{\mathcal{Q}}_n(r;\varepsilon_0)$ of radius $r/\sqrt{n}$ about the ideal distribution.This modification is negligible exponentially, but suffices to establish uniform higher order asymptotics for the MSE of the kind

$$\max_{Q_n \in \tilde{\mathcal{Q}}_n(r;\varepsilon_0)} n\,\mathrm{MSE}(S_n, Q_n) = r^2 \sup \psi^2 + \mathrm{E}_{\mathrm{id}}\psi^2 + \tfrac{r}{\sqrt{n}}\, A_1 + \tfrac{1}{n}\, A_2 + \mathrm{o}(\tfrac{1}{n}),$$

where A_1, A_2 are constants depending on ψ and r. Moreover, we essentially characterize contaminations generating maximal MSE up to $\mathrm{o}(n^{-1})$. Our results are confirmed empirically by simulations as well as numerical evaluations of the risk. With the techniques used for the MSE, we determine higher order expressions for the risk based on over-/undershooting probabilities as in Huber (1968) and Rieder (1980), respectively.

In the symmetric case, we find the second order optimal scores again of Hampel form, but to an $\mathrm{O}(n^{-1/2})$-smaller clipping height c than in first order asymptotics. This smaller c improves MSE only by $\mathrm{O}(n^{-1})$. For the case of unknown contamination radius we generalize the minimax inefficiency introduced in Rieder et al. (2001) to our second order setup. Among all risk maximizing contaminations we determine a "most innocent" one. This way we quantify the "limits of detectability"in Huber (1997)'s definition for the purposes of robustness.

*..

Contents

1 Motivation/introduction

1.1 Setup: one-dimensional location

This paper deals with the one–dimensional location model, i.e.

$$X_i = \theta + v_i, \qquad v_i \overset{\text{i.i.d.}}{\sim} F, \qquad P_\theta = \mathcal{L}(X_i) \tag{1.1}$$

for some ideal distribution F with finite Fisher-Information of location $\mathcal{I}(F)$, i.e.

$$\Lambda_f = -\dot{f}/f \in L_2(F), \qquad \mathcal{I}(F) = \mathrm{E}[\Lambda_f^2] < \infty \tag{1.2}$$

We also assume that Λ_f is increasing. By translation equivariance, we may restrict ourselves to $\theta_0 = 0$ which will be suppressed in the notation.

Following Rieder (1994), we may define the set of *influence curves* (IC's) Ψ for the estimation of θ as

$$\Psi := \{\psi \in L_2(F) \,|\, \mathrm{E}[\psi] = 0, \quad \mathrm{E}[\psi\Lambda_f] = 1\}, \tag{1.3}$$

where both expectations are evaluated under F. As class of estimators we consider *asymptotically linear estimators* (ALE's), i.e. estimators $S_n = S_n(X_1, \ldots, X_n)$ with the property

$$\sqrt{n}\; S_n = \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n \psi(X_i) + \mathrm{o}_{F^n}(n^0) \tag{1.4}$$

1.2 Shrinking neighborhoods

We will consider the maximal mean squared error (MSE) on neighborhoods of this ideal model. To avoid dominance of the bias for increasing number of observations, we follow Rieder (1994), working in the setup of shrinking neighborhoods. For this paper we consider contamination neighborhoods, i.e. the set $\mathcal{Q}_n(r)$ of distributions

$$\mathcal{L}_\theta^{\rm real}(X_1,\ldots,X_n) = Q_n = \bigotimes_{i=1}^n [(1-\tfrac{r_n}{\sqrt{n}})F + \tfrac{r_n}{\sqrt{n}}\,P_{n,i}^{\rm di}] \tag{1.5}$$

with $r_n = \min(r,\sqrt{n})$, $r>0$ the contamination radius and $P_{n,i}^{\rm di} \in \mathcal{M}_1(\mathbb{B})$ arbitrary, uncontrollable contaminating distributions. As usual, we interpret Q_n as the distribution of the vector $(X_i)_{i\le n}$ with components

$$X_i := (1-U_i)X_i^{\rm id} + U_i X_i^{\rm di}, \qquad i=1,\ldots,n \tag{1.6}$$

for $X_i^{\rm id}$, U_i, $X_i^{\rm di}$ stochastically independent, $X_i^{\rm id} \stackrel{\rm i.i.d.}{\sim} F$, $U_i \stackrel{\rm i.i.d.}{\sim} {\rm Bin}(1,r/\sqrt{n})$, and $(X_i^{\rm di}) \sim P_n^{\rm di}$ for some arbitrary $P_n^{\rm di} \in \mathcal{M}_1(\mathbb{B}^n)$.

1.3 First order optimality

For a sequence of estimators S_n, consider the following asymptotic (modified) maximal mean squared error on $\mathcal{Q}_n$

$$\tilde R(S_n,r) := \lim_{t\to\infty}\lim_{n\to\infty}\sup_{Q_n\in\mathcal{Q}_n(r)}\int \min\{t,\, n\,|S_n-\theta_0|^2\}\,dQ_n \tag{1.7}$$

In Rieder (1994), it is shown that in the general p-dimensional L_2-differentiable model, with scores Λ_θ and Fisher-Information $\mathcal{I}_\theta$ (suppressing the dependency upon θ as usual) a (suitably constructed) ALE S_n with IC ψ has risk

$$\tilde R(S_n,r) = r^2\sup|\psi|^2 + {\rm E}_{\rm id}\,|\psi|^2 \tag{1.8}$$

In Theorem 5.5.7 (ibid.), together with its preceding remarks, it is proved that, for given $r\ge 0$, among all such ALEs, any (suitably constructed) ALE with IC η_{b_0} minimizes $\tilde R(\,\cdot\,,r)$ where η_{b_0} is of Hampel form

$$\eta_{b_0} = Y\min\{1, b_0/|Y|\}, \qquad Y = A\Lambda - a \tag{1.9}$$

for some $A\in\mathbb{R}^{p\times p}$, $a\in\mathbb{R}^p$ such that η_{b_0} is an IC, and b_0 solving ${\rm E}(|Y|-b_0)_+ = r^2 b_0$. In our context, for Lagrange multipliers z and A such that $\eta_{b_0}=\eta_{c_0}\in\Psi$, we get that

$$\eta_{c_0} = A(\Lambda_f - z)\min\{1, c_0/|\Lambda_f - z|\} \tag{1.10}$$

$$c_0 \qquad \text{s.t.}\quad {\rm E}[(|\Lambda_f - z| - c_0)_+] = r^2 c_0 \tag{1.11}$$

1.4 Open issues in this setup

Being bound to first order asymptotics, so far these results do not come along with an indication for the speed of the convergence; it is not clear to what degree radius r, sample size n and clipping height b affect this approximation. The theorem only characterizes the optimal expansion in terms of ICs. Finally, modification (1.7) of the MSE, which is common in as. statistics, confer Le Cam (1986), Rieder (1994), Bickel et al. (1998), van der Vaart (1998), and which forces the integrals to converge under weak convergence, appears somewhat ad hoc. One would perhaps prefer a modification that is statistically motivated.

1.5 M-estimators for location

As estimators to achieve (1.4) for a given IC ψ, we consider M-estimators. More specifically we require ψ to be monotone and bounded and write $\psi_t(\,\cdot\,)$ for $\psi(\,\cdot\,-t)$. For technical reasons we assume that the law of $\psi_t(X^{\rm id})$ has non-trivial absolutely continuous component uniformly in t —compare condition (C)/(C') later; in particular the set D_t of discontinuities of the c.d.f. of $\psi_t(X^{\rm id})$ has to carry less mass than 1 uniformly:

$$p_D := \sup_t\, P^{\rm id}(D_t) < 1 \tag{1.12}$$

Following the notation in Huber (1981, pp. 46), let

$$S_n^* := \sup\Big\{\, t \,|\, \sum_{i\le n} \psi_t(x_i) > 0\Big\}, \qquad S_n^{**} := \inf\Big\{\, t \,|\, \sum_{i\le n} \psi_t(x_i) < 0\Big\} \tag{1.13}$$

and S_n be any estimator satisfying $S_n^* \le S_n \le S_n^{**}$. By monotonicity of ψ, we get

$$\Pr\{S_n^* < t\} = \Pr\Big\{\sum_{i\le n} \psi_t(x_i) \le 0\Big\}, \qquad \Pr\{S_n^{**} < t\} = \Pr\Big\{\sum_{i\le n} \psi_t(x_i) < 0\Big\} \tag{1.14}$$

in the continuity points t of the LHS. The next lemma, an immediate consequence of Hall (1992, Theorem 2.3), shows that we may ignore the event $S_n^* \ne S_n^{**}$ if we are interested in statements valid up to $\mathrm{o}(1/n)$.

Lemma 1.1 *Under* (1.12)*,* $\Pr(S_n^* \ne S_n^{**}) = \mathrm{O}(\exp(-\gamma n))$ *for some* $\gamma > 0$.

Remark 1.2 If $\bigcup_t D_t = \{\pm c\}$ for some $c > 0$, $\Pr(S_n^* \ne S_n^{**}) = 0$ for n odd.

Remark 1.3 In principle, the arguments used in our paper are not confined to the location case. In fact, we crucially use monotony of the scores function/IC. To cover multivariate M-estimators or M-estimators with a non-monotone IC, an approach local to a $\sqrt{n}$–consistent starting esimtator seems to be more appropriate. We have not yet worked this out, however.

1.6 Organization of this paper and description of the results

In this paper, we will provide answers to some of the open questions mentioned in subsection 1.4: In a closer inspection of simulations, M. Kohl found out that larger inaccuracies of (first order) asymptotics only occurred when there were extraneous sample situations where more than half the sample size stemmed from a contamination, which made him conjecture that excluding such samples, asymptotics might then prove useful even for very small samples. In fact this gives a convenient modification of the contamination neighborhood system based on Donoho and Huber (1983)'s finite sample breakdown point ε_0 for S_n. This modification on the one hand is asymptotically negligible hence does not affect the results of subsection 1.2, but on the other hand enforces the unmodified MSE to converge along with weak convergence. We will start with presenting this modification in section 2. In section 3, we then present the central theoretical result, Theorem 3.6. This result is of the following form

$$\sup_{Q_n\in\tilde{\mathcal{Q}}_n(r;\varepsilon_0)} n\,\mathrm{MSE}(S_n,Q_n) = r^2\sup|\psi|^2+\mathrm{E}\,\psi^2+\tfrac{r}{\sqrt{n}}\,A_1+\tfrac{1}{n}\,A_2+\mathrm{o}(\tfrac{1}{n}) \qquad (1.15)$$

Here S_n is an M-estimator to IC ψ, and A_1, A_2 are polynomials in the contamination radius r, in $b=\sup|\psi|$, and in the moment functions $t\mapsto \mathrm{E}\,\psi_t^l$, $l=1,\ldots,4$ and their derivatives evaluated in $t=0$. We recognize a uniform higher order as. expansion for the MSE along the modified neighborhood and that the speed of the convergence to the first order as. value is one order faster in the ideal model.

Notation 1.4 For indices we start counting with 0, so that terms of first-order asymptotics have an index 0, second-order ones a 1 and so on. Also we abbreviate first-order, second-order and third-order by f-o, s-o, t-o respectively, and we write f-o-o, s-o-o, and t-o-o for first, second, and third-order asymptotically optimal respectively.

As to the correctness of our main result, we give a number of cross checks and comments on this result in section 4. That these results are already relevant for small sample sizes is shown by a simulation study which is presented in section 5 as to its design and results. This section will also contain numerical results obtained with an adopted convolution algorithm taken from Kohl et al. (2004). In particular, our main result compares fairly well with results obtainable in the fixed-neighborhood setup, compare Fraiman et al. (2001), with the advantage of explicit expressions instead of numerical solutions.

Some ramifications of Theorem 3.6 are presented in section 6: With a slight (further) restriction of the neighborhood system, we make our main result available in the case that the central distribution has tails decaying at a polynomial rate in Proposition 6.1. As is shown in Proposition 6.2, polynomial tails are essentially necessary for a finite MSE at least in the ideal model. The sufficient condition for a sequence of contaminations to achieve maximal risk from Theorem 3.6 is shown to be almost necessary in Proposition 6.3.

As examples for the wide application range of the techniques used to prove this theorem, we determine higher order expansions for bias and variance separately in Proposition 6.4. In Theorem 6.5, we take up the risk consisting in certain over-and undershooting probabilities used in Huber (1968) to determine a finite sample minimax estimator of location. By means of a s-o expansion, we refine the corresponding f-o translation by Rieder (1980), providing a closer link to finite sample optimality.

Some consequences of Theorem 3.6 are discussed in section 7. In subsection 7.1, we show for F symmetric in θ, the f-o optimality of Hampel-type ICs of form (1.10) persists if we account for the A_1 term in (1.15). Hence, in this sense, Pfanzagl (1979)'s catchword "*First order efficiency implies second order efficiency*" survives (at least partially) when passing to neighborhoods around the ideal (symmetric) model. We even may determine the s-o-o clipping height $c_1 = c_1(r,n)$ which in fact is slightly lower ($\mathrm{O}(n^{-1/2})$) than the f-o-o $c_0 = c_0(r)$ determined according to (1.11). Passing to c_1, as. MSE can at most be improved by $\mathrm{O}(n^{-1})$. So in fact we only retain the optimal class, not the actual optimal estimator from f-o optimality.

A (partial) explanation for the good, respectively excellent behaviour of f-o-o, s-o-o and t-o-o procedures as to numerically exact finite maximal MSE, we present an argument based on a functional implicit function theorem in section 7.2.

For decisions upon the procedure to take, only relative risk is relevant which is discussed in some detail in subsection 7.3. We then proceed to obtain a s-o variant of the minimax radius introduced and determined in Rieder et al. (2001): In the situation where the radius is unknown within a range (r_l, r_u), a radius r_0 is determined such that the (f-o) maximal inefficiency $\bar\rho(r')$ defined in (7.17) is minimized in $r' = r_0$. We translate this to the s-o setup in section 7.4; the s-o results in the Gaussian location model show that neither $c_1(r_1,\cdot)$, nor s-o minimax radius $r_1(\cdot)$ vary much in n and that for all n, s-o minimax inefficiency is always smaller than the corresponding f-o one.

We also get a deeper insight to the question which contaminations are (already) dangerous; in subsection 7.5, we determine a most innocent appearing least favorable contamination which is shown to form a saddlepoint together with the f-o (s-o) optimal M-estimator. It appears to be innocent, as it produces only "outliers" which are hardest to detect in some sense specified in this section.

In the following section 8, we present proofs to the theorems and propositions of this paper. These contain rather tedious Taylor expansions where we need the help of a symbolic Algebra program like `MAPLE`. To ease readability, we therefore start the proof of the main theorem with an outline of the essential steps. Some auxiliary results needed in the proofs are provided in an appendix in section 9.

For the interested reader we have set up a web-page to this article under
`http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/mest.html`
On this page, additional tables and figures, the `MAPLE` script to generate the expansions, and the `R`-script to calculate numerically exact MSE are available for download.

1.7 Deferred problems

The question of what construction principle to take will be discussed in a subsequent paper, Ruckdeschel (2005b), where we will present an analogue to the main theorem of this paper for the One-Step-construction principle.

2 Modification of the shrinking neighborhood setup

The key property in the shrinking-neighborhood setup is the LAN-property[1] in the sense of Hájek and LeCam. This property is generally available in L_2-differentiable models, c.f. Rieder (1994, Thm. 2.3.5). This property together with LeCam's third Lemma —c.f. Corollary 2.2.6 ibid.— implies uniform weak convergence of any (suitably constructed) ALE to a bounded IC on a representative subclass of the system of neighboring distributions $\mathcal{Q}_n$ — those distributions induced by simple perturbations $Q_n(\zeta, t)$, confer p. 126 (ibid.).
This weak convergence however does not entail convergence of the risk for an unbounded loss function in general, as we show in the following example:

2.1 Convergence failure of the MSE for M-estimators to bounded scores

Proposition 2.1 *Let $\mathcal{P}$ be the location model from* (1.1). *Let ψ be an isotone influence curve with* $\sup|\psi| = b < \infty$ *which is Lipschitz bounded. Let S_n be an M-Estimator according to* (1.13) *that is uniformly consistent on $\mathcal{Q}_n$. Then for sample size n, for each $\theta \in \mathbb{R}$ and each $K_n \uparrow \infty$ there is a sequence $x_n \in \mathbb{R}$ such that with $Q_n = [(1 - \frac{r_n}{\sqrt{n}})P_\theta + \frac{r_n}{\sqrt{n}}\,\mathrm{I}_{\{x_n\}}]^n$*

$$n\,\mathrm{MSE}(S_n, Q_n) > K_n \tag{2.1}$$

although, with $T(Q)$ the zero of $t \mapsto \int \psi_t\,dQ$, it holds that uniformly in $\mathcal{Q}_n$,

$$\sqrt{n}\,(S_n - T(Q_n)) \circ Q_n \xrightarrow{\mathrm{w}} \mathcal{N}(0, \mathrm{E}_F[\psi_0^2]) \tag{2.2}$$

2.2 Modification of the shrinking neighborhood setup

The proof of proposition 2.1, suggests the following modification for finite n: Only such realizations of $U_1, \ldots, U_n$ are permitted, where $\sum U_i < n/2$ —the case $\sum U_i = n/2$ only occurs for even sample size and will not be considered here.
More precisely, accounting for non-symmetric ψ, we introduce

$$\check{b} := \inf\psi, \qquad \hat{b} = \sup\psi, \qquad \bar{b} := \tfrac{1}{2}(\hat{b} - \check{b}), \qquad \delta_0 := \tfrac{|(-\check{b}) - \hat{b}|}{\min((-\check{b}), \hat{b})} \tag{2.3}$$

and recall that in our situation, both the functional (Huber, 1981, (2.39),(2.40)) and the finite sample (ε-contamination) breakdown point (Donoho and Huber, 1983, section 2.2) of T respectively S_n are

$$\varepsilon_0 = 1/(2 + \delta_0) \tag{2.4}$$

With these expressions, our modifiation amounts to considering the neighborhood system $\tilde{\mathcal{Q}}_n(r; \varepsilon_0)$ of conditional distributions

$$Q_n = \mathcal{L}\Big\{[(1 - U_i)X_i^{\mathrm{id}} + U_i X_i^{\mathrm{di}}]_i \,\Big|\, \sum U_i \le \ulcorner \varepsilon_0 n \urcorner - 1\Big\} \tag{2.5}$$

[1] for local as. normality

This restriction hence combines a restriction to the marginals $\mathcal{L}(X_i^{\rm real})$ which are "close" to $\mathcal{L}(X_i^{\rm id})$ for each i as well as a sample-wise restriction.
Correspondingly, we will consider the asymptotics of

$$R_n(S_n, r; \varepsilon_0) := \sup_{Q_n \in \tilde{\mathcal{Q}}_n(r;\varepsilon_0)} n \int |S_n - \theta_0|^2 \, dQ_n \tag{2.6}$$

2.3 Asymptotic negligibility of this modification

The effect of this modification is negligible asymptotically: By the Hoeffding bound (9.1),

$$P(\sum U_i \geq n\varepsilon_0) \quad \leq \quad \exp(-2n(\varepsilon_0 - r/\sqrt{n}\,)^2) \tag{2.7}$$

which decays exponentially fast. Thus all results on convergence in law of the shrinking neighborhood setup are not affected when passing from $\mathcal{Q}_n(r)$ to $\tilde{\mathcal{Q}}_n(r; \varepsilon_0)$.

Remark 2.2 (a) Replacing $r/\sqrt{n}$ by the fixed radius ε, asymptotic negligibility continues to hold, as long as $\varepsilon < \varepsilon_0$.
(b) This concept of thinning out the neighborhoods according to the finite finite sample breakdown point easily generalizes to other setups; this has been spelt out in some detail in Ruckdeschel (2005a).

3 Main Theorem

Before the statement of the theorem, we introduce some auxiliary terms.

3.1 Notation

To $\psi : \mathbb{R} \to \mathbb{R}$ monotone let $\psi_t(x) := \psi(x - t)$ and define the following functions

$$L(t) := \mathrm{E}\, \psi(X - t), \qquad V(t)^2 := \mathrm{Var}\, \psi(X - t), \tag{3.1}$$

$$\rho(t) := \mathrm{E}[(\psi(X - t) - L(t))^3]/V(t)^3, \qquad \kappa(t) := \mathrm{E}[(\psi(X - t) - L(t))^4]/V(t)^4 - 3 \tag{3.2}$$

Let $\check{y}_n$ and $\hat{y}_n$ sequences in $\mathbb{R}$ such that for some $\gamma > 1$

$$\psi(\check{y}_n) = \inf \psi + \mathrm{o}(\tfrac{1}{n^\gamma}), \qquad \psi(\hat{y}_n) = \sup \psi + \mathrm{o}(\tfrac{1}{n^\gamma}) \tag{3.3}$$

To state our main theorem, we need the following notation:
For $H \in \mathcal{M}_1(\mathbb{B}^n)$ and an ordered set of indices $I = (1 \leq i_1 < \ldots < i_k \leq n)$ denote H_I the marginal of H with respect to I.

Definition 3.1 *Consider three sequences c_n, d_n, and κ_n in $\mathbb{R}$, in $(0, \infty)$, and in $\{1, \ldots, n\}$, respectively. We say that the sequence $(H^{(n)}) \subset \mathcal{M}_1(\mathbb{B}^n)$ is κ_n–*concentrated left [right] of c_n up to $\mathrm{o}(d_n)$, *if for each sequence of ordered sets I_n of cardinality $i_n \leq \kappa_n$*

$$1 - H_{I_n}^{(n)}\big((-\infty; c_n]^{i_n}\big) = \mathrm{o}(d_n) \qquad \Big[1 - H_{I_n}^{(n)}\big((c_n, \infty)^{i_n}\big) = \mathrm{o}(d_n)\Big] \tag{3.4}$$

3.2 Assumptions

The following assumptions will be needed for the main result of this paper:

(bmi) $\sup \|\psi\| = b < \infty\,,\ \psi$ monotone, $\psi \in \Psi$

(D) For some $\delta \in (0,1]\,,\ L\,,\ V\,,\ \rho$, and κ allow the expansions

$$L(t) \;=\; l_1 t + \tfrac{1}{2} l_2\, t^2 + \tfrac{1}{6} l_3\, t^3 + \mathrm{O}(t^{3+\delta}) \tag{3.5}$$
$$V(t) \;=\; v_0(1 + \tilde v_1\, t + \tfrac{1}{2}\tilde v_2\, t^2) + \mathrm{O}(t^{2+\delta}) \tag{3.6}$$
$$\rho(t) \;=\; \rho_0 + \rho_1\, t + \mathrm{O}(t^{1+\delta}) \tag{3.7}$$
$$\kappa(t) \;=\; \kappa_0 + \mathrm{O}(t^{\delta}) \tag{3.8}$$

(Vb) $V(t) = \mathrm{O}(|t|^{-(1+\delta)})$ for $|t| \to \infty$ and some $\delta \in (0,1]$

(C) Let f_t be the characteristic function of $\psi_t(X^{\mathrm{id}})$; then

$$\lim_{t_0 \to 0} \limsup_{s \to \infty} \sup_{|t| \le t_0} |f_t(s)| < 1 \tag{3.9}$$

Condition (C) is a local uniform Cramér condition; it is implied by

Lemma 3.2 *Assume $\mathcal{L}(\psi(X^{\mathrm{id}}))$ has a nontrivial absolute continuous part and that ψ is continuous. Then (C) is fulfilled.*

Remark 3.3 (a) By condition (bmi) —as $\psi \in \Psi$—, $l_1 = -1$.

(b) Condition (C) is not fulfilled for the median, as its influence curve just takes the values $-b, b$ F-a.e. A direct proof for an analogue to Theorem 3.6 is possible, however, and given in Ruckdeschel (2005a).

(c) If one is content with an expansion of the MSE up to order $\mathrm{o}(n^{-1/2})$, we may drop (3.8) and use the following weakened assumptions

(D') For some $\delta \in (0,1]\,,\ L\,,\ V$, and ρ allow the expansions

$$L(t) \;=\; l_1 t + l_2/2\, t^2 + \mathrm{O}(t^{2+\delta}), \tag{3.10}$$
$$V(t) \;=\; v_0(1 + \tilde v_1\, t) + \mathrm{O}(t^{1+\delta}) \tag{3.11}$$
$$\rho(t) \;=\; \rho_0 + \mathrm{O}(t^{\delta}) \tag{3.12}$$

(C') "Uniformly" for t around $t = 0\,,\ \mathcal{L}\big(\psi_t(X^{\mathrm{id}})\big)$ is not a lattice distribution, that is, there exist $t_0 > 0\,,\ s_0 > 0$ such that for all $s_1 > s_0$

$$\hat f_{s_0,t_0}(s_1) := \sup_{s_0 \le s \le s_1} \sup_{|t| \le t_0} |f_t(s)| < 1 \tag{3.13}$$

Note that (C) implies (C'), but contrary to (C), in (C') the case $\sup_{s_1} \hat f_{s_0,t_0}(s_1) = 1$ for all $s_0 > 0$ and all $t_0 > 0$ is allowed.

3.3 Illustration

We specialize assumptions (bmi) to (C) for $F = \mathcal{N}(0,1)$ and $\psi \in \Psi$ of form (1.10).

Proposition 3.4 *For $F = \mathcal{N}(0,1)$ and $\psi = \eta_c$ an IC to some $c \in (0,\infty)$ of Hampel–form $\eta_c = A_c(x \min\{1, \frac{c}{|x|}\}$, assumptions (bmi) to (C) are in force; in particular the bound in (Vb) holds even exponentially.*

Remark 3.5 For η_c to be an IC, $A_c = (2\Phi(c)-1)^{-1}$. As to the terms from (D) we get, with $\Phi(x)$ the c.d.f. of $\mathcal{N}(0,1)$ and $\varphi(x)$ its density

$$l_2 = 0, \qquad \tilde v_1 = 0, \qquad \rho_0 = 0 \tag{3.14}$$

For $c \in (0,\infty)$, we get

$$l_3 = 2c\varphi(c)/(2\Phi(c)-1) \tag{3.15}$$
$$v_0^2 = 2b^2(1-\Phi(c)) + A_c(1-2b\varphi(c)) \tag{3.16}$$
$$\tilde v_2 = \frac{6\Phi(c) - 4\Phi(c)^2 - 2 - 2c\varphi(c)}{2c^2(1-\Phi(c)) + 2\Phi(c) - 1 - 2c\varphi(c)} \tag{3.17}$$
$$\rho_1 = \frac{3A_c^3\,(1 - 2\Phi(c) + 2c\varphi(c))}{v_0^3} + 3v_0^{-1} \tag{3.18}$$
$$\kappa_0 = \frac{2c^4\,(1-\Phi(c)) - 2c(c^2+3)\varphi(c) + 3(2\Phi(c)-1)}{[2c^2\,(1-\Phi(c)) + 2\Phi(c) - 1 - 2c\varphi(c)]^2} - 3 \tag{3.19}$$

For $c \downarrow 0$, $\quad l_3 = 1, \quad v_0^2 = \frac{\pi}{2}, \quad \tilde v_2 = -\frac{2}{\pi}, \quad \rho_1 = 2\sqrt{\frac{2}{\pi}}, \quad \kappa_0 = -2$, and, formally, for $c \uparrow \infty$, $\quad l_3 = 0, \quad v_0 = 1, \quad \tilde v_2 = 0, \quad \rho_1 = 0, \quad \kappa_0 = 0$.

3.4 Statement of the main theorem

Theorem 3.6 (Main Theorem) *In the location model* (1.1) *with* (1.2) *assume (bmi) to (C) from section 3.2. Then for sample size n,*

(a) *the following expansion of the maximal MSE of an an M-estimator S_n to scores-function ψ holds*

$$R_n(S_n, r, \varepsilon_0) \;=\; r^2b^2 + {v_0}^2 + \tfrac{r}{\sqrt{n}}\,A_1 + \tfrac{1}{n}\,A_2 + \mathrm{o}(n^{-1}) \tag{3.20}$$

with

$$A_1 = {v_0}^2\Big(\pm(4\,\tilde v_1 + 3\,l_2\,)b + 1\Big) + b^2 + [2\,b^2 \pm l_2\,b^3\,]\,r^2 \tag{3.21}$$

$$\begin{aligned} A_2 = {} & {v_0}^3\Big((l_2 + 2\,\tilde v_1\,)\rho_0 + \tfrac{2}{3}\,\rho_1\Big) + {v_0}^4\,(3\,\tilde v_2 + \tfrac{15}{4}\,{l_2}^2 + l_3 + 9\,\tilde v_1^2 + 12\,\tilde v_1\,l_2\,) + \\ & +[\,{v_0}^2\Big((3\,\tilde v_2 + 3\,\tilde v_1^2 + \tfrac{15}{2}\,{l_2}^2 + 2\,l_3 + 12\,\tilde v_1\,l_2\,)b^2 + 1 \pm (8\,\tilde v_1 + 6\,l_2\,)\,b\Big) + \\ & \pm 3\,l_2\,b^3 + 5\,b^2\,]\,r^2 + \Big((\tfrac{5}{4}\,{l_2}^2 + \tfrac{1}{3}\,l_3\,)b^4 \pm 3\,l_2\,b^3 + 3\,b^2\Big)\,r^4 \end{aligned} \tag{3.22}$$

and we are in the $-[+]$-case depending on whether (3.23) *or* (3.24) *below applies.*

(b) *let* $P_n^{\rm di} := \bigotimes_{i=1}^n P_{n,i}^{\rm di}$ *be contaminating measures for* (1.5). *Then* Q_n *with* $P_n^{\rm di}$ *as contaminating measures achieves the maximal risk in* (3.20) *if for* $k_1 > 1$ *and* $k_2 > 2 \vee (\frac{3}{2} + \frac{3}{2\delta})$ *with* δ *from (Vb) and* $K_1(n) = \ulcorner k_1 r \sqrt{n} \urcorner$ *either*

$$(P_n^{\rm di}) \textit{ is } K_1(n)\textit{–concentrated left of } \check y_n - b\sqrt{k_2 \log(n)/n} \textit{ up to } {\rm o}(n^{-1}) \tag{3.23}$$

or

$$(P_n^{\rm di}) \textit{ is } K_1(n)\textit{–concentrated right of } \hat y_n + b\sqrt{k_2 \log(n)/n} \textit{ up to } {\rm o}(n^{-1}) \tag{3.24}$$

More precisely, if $\sup\psi\;<\;[>]\;-\inf\psi$, *the maximal MSE is achieved by contaminations according to* (3.23) *[*(3.24)*]. In case* $\sup\psi = -\inf\psi$, (3.23) *[*(3.24)*] applies if*

$$\tilde v_1 \;>\; [<] - \tfrac{l_2}{4}\Big(\tfrac{b^2}{v_0^2}(r^2+3)(1+\tfrac{r}{\sqrt{n}}-\tfrac{2r^2}{n}) + 3(1-\tfrac{b^2}{v_0^2})\Big) \tag{3.25}$$

If $\sup\psi = -\inf\psi$ *and there is "*$=$*" in* (3.25), (3.23) *and* (3.24) *generate the same risk up to order* ${\rm o}(n^{-1})$.

Remark 3.7 (a) Curiously, although being of corresponding order, no ρ_0 $[\kappa_0]$-term shows up in the correction term A_1 $[A_2]$, which is probably due to the special loss function. We thus conjecture that we may dispense of condition (C') for s-o asymptotics for the MSE.

(b) As announced in the introduction, for $r=0$, we get an approximation that is one order faster than under contamination.

(c) Let Q_n^0 be any distribution in $\tilde{\mathcal{Q}}_n$ attaining maximal risk in Theorem 3.6. Under symmetry or more specifically if $l_2 = v_1 = \rho_0 = 0$, (3.20) becomes

$$n\;{\rm E}_{Q_n^0}[S_n^2] = \big(r^2b^2 + {v_0}^2\big)\left(1 + \tfrac{r}{\sqrt{n}}\right) + \tfrac{r}{\sqrt{n}}\,\big(b^2(1+r^2)\big) + {\rm O}(n^{-1}) \tag{3.26}$$

Thus under symmetry and for large enough n, **the maximal MSE on $\tilde{\mathcal{Q}}_n$ is always underestimated by f-o asymptotics!**

(d) In the ideal Gaussian location model (i.e. $r=0$), plugging in the (limiting) results for $c=0$ from section 3.3, the RHS of (3.26) becomes

$$\frac{\pi}{2}\left(1 + \frac{1}{n}(\frac{\pi}{2} - \frac{5}{3})\right) + {\rm o}(n^{-1}) \doteq 1.5708(1 - \tfrac{0.0958}{n}) + {\rm o}(n^{-1}) \tag{3.27}$$

suggesting an overestimation of the risk by the f-o asymptotics. This is to be compared to the result for the median for odd sample size from Ruckdeschel (2005a):

$$n\;{\rm E}_{F^n}[{\rm Med}_n^2] = \frac{\pi}{2}\left(1+\frac{1}{n}(\frac{\pi}{2}-2)\right)+{\rm o}(n^{-1}) \doteq 1.5708(1-\tfrac{0.4292}{n})+{\rm o}(n^{-1}) \tag{3.28}$$

Hence we indeed overestimate the risk by the f-o asymptotics. The difference of $\frac{\pi}{6\,n}$ is due to the failure of condition (C).

3.5 Cross-checks

3.5.1 Check with results by Fraiman et al.

In the symmetric case, the first cross check comes with the as. formula for variance $\mathrm{asVar}(\psi)$ and (maximal) bias $\mathrm{asBias}(\psi)$ as to be found in Fraiman et al. (2001), where we have to identify $\varepsilon = r/\sqrt{n}$.

$$\mathrm{asBias}(\psi)/\sqrt{n} := B(\psi) = \{\beta \,|\, (1-\varepsilon)\int \psi_\beta \, dF + \varepsilon b = 0\} \tag{3.29}$$

$$\mathrm{asVar}(\psi) := \frac{(1-\varepsilon)\int \psi^2_{B(\psi)}\, dF + \varepsilon b^2}{(1-\varepsilon)^2 (\int \dot\psi_{B(\psi)}\, dF)^2}, \tag{3.30}$$

Assuming that $\int \dot\psi_{B(\psi)}\, dF = L'(B(\psi))$ and using that

$$\begin{aligned}
\int \psi_{B(\psi)}\, dF &= L(B(\psi)) = -B(\psi) + \mathrm{o}(B^2)) \\
\int \psi^2_{B(\psi)}\, dF &= V(B(\psi))^2 + L(B(\psi))^2 = v_0^2(1+\mathrm{o}(B)) \\
L'(B(\psi))^2 &= -1 + \mathrm{o}(B)
\end{aligned}$$

we get that

$$\mathrm{asBias}(\psi) = \sqrt{n}\, b\varepsilon(1+\varepsilon+\mathrm{o}(\varepsilon)) = rb(1+\tfrac{r}{\sqrt{n}} + \mathrm{o}(n^{-1/2})) \tag{3.31}$$

$$\mathrm{asVar}(\psi) = (1+\varepsilon)v_0^2 + \varepsilon b + \mathrm{o}(\varepsilon) = v_0^2 + \tfrac{r}{\sqrt{n}}(v_0^2 + b) + \mathrm{o}(n^{-1/2}) \tag{3.32}$$

and hence —in accordance with formula (3.20)—

$$\mathrm{asMSE}(\psi) = (v_0^2 + r^2 b^2)(1+\tfrac{r}{\sqrt{n}}) + \tfrac{r}{\sqrt{n}} b^2(1+r^2) + \mathrm{o}(n^{-1/2}) \tag{3.33}$$

3.5.2 Check with second order asymptotics for the median

The second check comes with the s-o asymptotics for the median from Ruckdeschel (2005a). To that end we assume that with $f_0 > 0$ and some $\delta \in (0,1]$,

$$f(t) = f_0 + f_1 t + \mathrm{O}(t^{1+\delta}) \tag{3.34}$$

As for the median, $\psi_{\mathrm{Med}} = \frac{1}{2\,f_0}\,\mathrm{sign}(x)$, we have $v_0 = b = \frac{1}{2f_0}$ and $\varepsilon_0 = 1/2$. For the moment we ignore the fact, that condition (C) —resp. (C')— is not fulfilled. Easy calculations give

$$l_2 = -f_1/f_0, \qquad \tilde v_1 = 0, \qquad \rho_0 = 0 \tag{3.35}$$

so that with our formula (3.20) we obtain for odd sample size n

$$R_n(\psi_{\mathrm{Med}_n}, r, \tfrac{1}{2}) = \frac{1}{4f_0^2}\Big((1+r^2)\left[1+\tfrac{2r}{\sqrt{n}}\right] - \tfrac{r}{\sqrt{n}}\frac{f_1}{2f_0^2}(r^2+3)\Big) + \mathrm{o}(n^{-1/2}) \tag{3.36}$$

in complete concordance with Ruckdeschel (2005a).

3.5.3 Check with third order asymptotics for the median

The third check takes up the second and compares t-o asymptotics to be obtained by (3.20) —again ignoring condition (C). We get

$$l_3 = -f_2/f_0, \qquad \tilde{v}_2 = -4f_0^2, \qquad \rho_1 = 4f_0 \tag{3.37}$$

and hence for odd sample size n, after some reordering

$$R_n(\psi_{\mathrm{Med}_n}, r, \tfrac{1}{2}) \stackrel{?}{=} \mathrm{o}(\tfrac{1}{n}) + \frac{1}{4f_0^2}\Big\{(1+r^2) + \tfrac{r}{\sqrt{n}}\Big(2(1+r^2) + \tfrac{r^2+3}{2}\frac{|f_1|}{f_0^2}\Big) + \\ + \tfrac{1}{n}\Big(\boxed{\tfrac{4}{3}} - 3 + 3r^2 + 3r^4 + \tfrac{3r^2(3+r^2)}{2}\frac{|f_1|}{f_0^2} - \tfrac{3+6r^2+r^4}{12}\frac{f_2}{f_0^3} + \tfrac{5(3+6r^2+r^4)}{16}\frac{f_1^2}{f_0^4}\Big)\Big\} \tag{3.38}$$

and it is just the framed term $\frac{4}{3}$, which is coming in as $\frac{2}{3}\rho_1 v_0$ from (3.22), which causes a difference to the result of Ruckdeschel (2005a), where we get the value 1 instead. This discrepancy, however, is in fact due to the failure of condition (C), because Theorem 9.3, which we need to prove (3.20), is not available in this case.

4 Relations to other approaches

4.1 Small sample asymptotics

Of course the idea of assessing the quality / speed of convergence of CLT-type arguments by means of higher order asymptotics is common in Mathematical Statistics, confer among others Ibragimov and Linnik (1971), Bhattacharya and Rao (1976), Pfanzagl (1985), Hall (1992), Barndorff-Nielsen and Cox (1994) and Taniguchi and Kakizawa (2000).

Asymptotic expansions of the moments of statistical estimators —like MSE in our case— have already been studied by Gusev (1976) and Pfaff (1977); both approaches, however, only consider the ideal model, and work with pointwise expansions of the likelihood.

Also the idea to improve convergence by means of saddlepoint techniques and conjugate densities, respectively, has been a large success in this context, confer Daniels (1954), Hampel (1974), Field and Ronchetti (1990).

Our approach is simpler in the sense that instead of approximating the c.d.f. or the density of our procedures on the whole range of arguments, we directly approximate our risk. Doing so, we do not run into problems of bad approximations in the tails of a distribution, because all that is interesting for our risk will occur within a (decreasing) compact; using saddlepoint techniques, we would have to solve the saddlepoint-equation for a grid of evaluation points t_i to get an accurate estimate for the density which makes the corresponding solution less explicite than ours.

Even more important, note that in view of Proposition 2.1, a highly accurate approximation of the distribution of the M-estimator would not suffice to enforce uniform convergence of the MSE, which was the reason for our modification of the neighborhoods (2.5). Also, contrary to "usual" small sample asymptotics, by our

approach no particular contamination has to be assumed right from the beginning but we rather identify a least favorable one within the proof.
In the setup of saddlepoint-approximations, one would probably apply Theorem 4.3 in Field and Ronchetti (1990) which at least covers the Hampel-type solutions. The pointwise formulation of assumption A4.2 therein,

A4.2 There is an open subset $U \subset \mathbb{R}$, such that
   (i) for each $\theta \in \mathbb{R}$, $F(U-\theta) = 1$
   (ii) $D\psi$, $D^2\psi$, $D^3\psi$ exist on U

however, seems a bit dangerous, as it allows for pathological ψ-functions defined similar to the Cantor distribution function (while F may be something like $\mathcal{N}(0,1)$), for which the interchange of differentiation and integration becomes awkward. As may be read off from (3.20), in the ideal model, as for the saddlepoint approach, we, too, get an expansion of order $1/n$, a fact, which is *not* due to symmetry of Λ and/or ψ! So in fact we get the same approximation quality as with the saddlepoint approach —indeed, by the Taylor-expansion step in section 8.4.8, we extract an argument to be expanded from the exponential, which also is an idea behind the saddlepoint approximation, confer Field and Ronchetti (1990, p. 26). On the other hand, even in the restricted neighborhoods of (2.5), it is not clear to the present author, if in general, the saddlepoint approximation holds uniformly in t, so it is not clear, whether an improved approximation for the density will result in a better approximation of the risk. A detailed empirical and numerical investigation of such questions is contained in Ruckdeschel and Kohl (2004).

4.2 Approach by Fraiman et al. (2001)

In Fraiman et al. (2001), the authors work in a similar setup, i.e. the one-dimensional location problem where the center distribution is $F_0 = \mathcal{N}(0,\sigma^2)$ and an M-estimator S_n to skew symmetric scores ψ is searched which minimizes the maximal risk on a neighborhood about F_0. Contrary to our approach, the authors work with convex contamination neighborhoods $\mathcal{V} = \mathcal{V}(F,\varepsilon)$ to a fixed radius ε.
There has been some discussion which approach —fixed or shrinking radius— is more appropriate, but for fixed sample size n, of course we may translate the fixed radius ε into our radius $r/\sqrt{n}$ and then compare the approximation quality of both approaches.
Fraiman et al. (2001) propose to use risks which are constructed by means of a positive function $g : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}_+$ of as. bias $b = b(G,\psi)$ —formula (3.29)— and as. variance $v^2 = v^2(G,\psi)$ —formula (3.30). The function g is assumed lower semicontinuous and symmetric in the first argument as well as isotone in each argument. The risk of an M-estimator to IC ψ is taken as the function

$$L_g(\psi) = \sup_{G\in\mathcal{V}} g(b(G,\psi), v(G,\psi)/n) \tag{4.1}$$

A mean squared error-type risk then is formed by $g(u,v) = u^2 + v$. It is not quite the MSE, as it employs the as. terms b and v and so their results may differ from ours. The crucial point is that to solve their optimization problem, they have to assume that besides bias, also variance is maximized (for their optimal $\hat{\psi}$) if we contaminate with a Dirac measure in ∞. According to this assumption, if we introduce $G_0 := (1-\varepsilon)F_0 + \varepsilon\,\mathrm{I}_{\{\infty\}}$, we have to find ψ minimizing

$$l_g(\psi) = g(b(G_0,\psi), v(G_0,\psi)/n) \tag{4.2}$$

Differently to the Hampel-type IC's the solutions to this problem are of form

$$\begin{aligned} \psi_{a,b,c,t}(x) &= \tilde{\psi}_{a,b,t}\big(x\min\{1,\tfrac{c}{|x|}\}\big), & (4.3)\\ \tilde{\psi}_{a,b,t}(x) &= a\tanh(tx) + b[x - t\tanh(tx)] & (4.4)\end{aligned}$$

but the "MSE"-optimal solutions are numerically quite close to corresponding Hampel-ICs ψ_H, for which the authors in turn show that always $L_g(\psi_H) = l_g(\psi_H)$. For an implementation of this optimization see the `R`-file `FYZ.R` available on the web-page.

A comparison

As a sort of benchmark for our results, we reproduce a comparison to be found in Ruckdeschel and Kohl (2004) —albeit in some more detail than in the cited reference: For the values of n and r from section 5, we determine the "MSE"-optimal $\hat{\psi}$ and a corresponding Hampel IC $\hat{\psi}_H$ which is then compared to the f-o-o and s-o-o IC derived in this paper. Within the class of Hampel-IC's, numerically, we also determine the t-o-o and the "exactly" optimal clipping-c, c_2 and $c_{\rm ex}$ respectively. We compare the resulting IC's as to their clipping-height and the corresponding (numerically exact) value of $R_n(S_n, r)$, denoted by MSE_n; the latter comparison is done by the terms $\mathrm{relMSE}_n^{\rm ex}(c_.)$, calculated as

$$\mathrm{relMSE}_n^{\rm ex}(c_.) = \Big(\frac{\mathrm{MSE}_n(c_.)}{\mathrm{MSE}_n(c_{\rm ex})} - 1\Big) \times 100\% \tag{4.5}$$

The results are displayed in Table 1. Also confer the function `allMSEs` in the `R`-file `asMSE.R` available on the web-page to this article.
For the numerical evaluation of the MSE, we use the techniques described in section 5.2. For $n = \infty$, we evaluate the corresponding f-o as. MSE for the IC to the corresponding values of c. As a cross-check, the clipping heights c_i, $i = 0, 1, 2$ are also determined for $n = 10^8$. In case of $c_{\rm FZY}$, for all finite n's the error tolerance used in `optimize` in `R` was 10^{-4}, while for $n = \infty$ it was 10^{-12}. For $c_{\rm ex}$ and $n = 10^8$, an optimization of the (numerically) exact MSE would have been too time-consuming and has been skipped for this reason. Also, for $n = 5$, the radius $r = 1.0$, corresponding to $\varepsilon = 0.447$, is not admitted for an optimization of (4.2) and thus no result is available in this case.

5 A simulation study and numerical evaluations

Before starting with the theoretical findings we summarize the results of a simulation study that actually lead us to the closer examination of the higher order expansions of the MSE.

Table 1: Optimal clipping heights and corresponding (numerically) exact MSE

r		$n=5$	$n=10$	$n=30$	$n=50$	$n=100$	$n=\infty$
0.1	c_0	1.948	1.948	1.948	1.948	1.948	1.948
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_0)$	8.679%	4.065%	1.340%	0.836%	0.448%	–
	c_1	1.394	1.484	1.611	1.663	1.724	1.948
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_1)$	0.833%	0.207%	0.027%	0.014%	0.010%	–
	c_2	1.309	1.428	1.585	1.644	1.713	1.948
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_2)$	0.332%	0.066%	0.008%	0.004%	0.006%	–
	c_{FZY}	1.368	1.370	1.610	1.668	1.756	1.939
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_{\mathrm{FZY}})$	0.658%	0.002%	0.026%	0.021%	0.031%	–
	c_{ex}	1.167	1.358	1.560	1.630	1.704	–
	$\mathrm{MSE}_n(c_{\mathrm{ex}})$	1.388	1.239	1.151	1.129	1.107	–
0.25	c_0	1.339	1.339	1.339	1.339	1.339	1.339
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_0)$	6.280%	3.681%	1.108%	0.656%	0.330%	–
	c_1	0.994	1.059	1.147	1.181	1.219	1.339
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_1)$	0.933%	0.415%	0.055%	0.023%	0.009%	–
	c_2	0.890	0.990	1.114	1.159	1.207	1.339
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_2)$	0.241%	0.104%	0.009%	0.002%	0.003%	–
	c_{FZY}	0.924	1.020	1.205	1.177	1.211	1.338
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_{\mathrm{FZY}})$	0.417%	0.215%	0.233%	0.018%	0.002%	–
	c_{ex}	0.783	0.921	1.092	1.140	1.205	–
	$\mathrm{MSE}_n(c_{\mathrm{ex}})$	2.225	1.705	1.438	1.381	1.330	–
0.5	c_0	0.862	0.862	0.862	0.862	0.862	0.862
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_0)$	2.930%	2.655%	0.792%	0.446%	0.218%	–
	c_1	0.650	0.690	0.746	0.767	0.790	0.862
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_1)$	0.756%	0.615%	0.087%	0.036%	0.013%	–
	c_2	0.547	0.620	0.712	0.744	0.777	0.862
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_2)$	0.230%	0.191%	0.015%	0.008%	0.003%	–
	c_{FZY}	0.539	0.632	0.716	0.749	0.782	0.866
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_{\mathrm{FZY}})$	0.200%	0.248%	0.021%	0.011%	0.008%	–
	c_{ex}	0.413	0.531	0.686	0.728	0.770	–
	$\mathrm{MSE}_n(c_{\mathrm{ex}})$	4.632	3.039	2.162	2.008	1.879	–
1.0	c_0	0.436	0.436	0.436	0.436	0.436	0.436
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_0)$	2.716%	3.132%	0.746%	0.348%	0.149%	–
	c_1	0.320	0.340	0.369	0.380	0.394	0.436
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_1)$	1.411%	1.610%	0.251%	0.076%	0.021%	–
	c_2	0.255	0.291	0.342	0.361	0.382	0.436
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_2)$	0.876%	0.999%	0.123%	0.027%	0.006%	–
	c_{FZY}	–	0.281	0.344	0.375	0.387	0.440
	$\mathrm{relMSE}_n^{\mathrm{ex}}(c_{\mathrm{FZY}})$	–	0.892%	0.132%	0.063%	0.012%	–
	c_{ex}	0.001	0.125	0.286	0.334	0.366	–
	$\mathrm{MSE}_n(c_{\mathrm{ex}})$	12.627	8.445	4.948	4.296	3.787	–

c	order	determined by	optimal among M-estimators
c_0	f-o-o	num. solution of (1.11)	to any IC
c_1	s-o-o	num. solution of (7.4)	in $\mathcal{S}_2$ (see section 7.1)
c_2	t-o-o	num. optimization of (3.20)	in $\mathcal{H}$ (see section 7.1)
c_{FZY}	—	num. optimization of (4.2)	to (4.4)-type ICs
c_{ex}	—	num. optimization of the (num.) exact MSE	in $\mathcal{H}$ (see section 7.1)

where (7.4) is the s-o analogue to (1.11), which is derived in Corollary 7.2. A more detailed description to this table is located on page 15.

5.1 Simulation design

Under `R 1.7.1`, we simulated $M = 10000$ runs of sample size $n = 5, 10, 30, 50, 100$ in the ideal location model $\mathcal{P} = \mathcal{N}(\theta, 1)$ at $\theta = 0$. In a contaminated situation, we used observations stemming from

$$Q_n = \mathcal{L}\{[(1-U_i)X_i^{\rm id} + U_i X_i^{\rm di}]_i \,\Big|\, \sum U_i \le \ulcorner n/2 \urcorner - 1\} \tag{5.1}$$

for $U_i \stackrel{\rm i.i.d.}{\sim} {\rm Bin}(1, r/\sqrt{n})$, $X_i^{\rm id} \stackrel{\rm i.i.d.}{\sim} \mathcal{N}(0,1)$, $X_i^{\rm di} \stackrel{\rm i.i.d.}{\sim} {\rm I}_{\{100\}}$ all stochastically independent and for contamination radii $r = 0.1, 0.25, 0.5, 1.0$.
As estimators we considered the median (with the mid-point variant for even sample size), and M-estimators to Hampel-type ICs η_c of form (1.10) with clipping heights $c = 0.5, 0.7, 1, 1.5, 2$ and $c_0(r)$, the f-o-o clipping height according to (1.11). All empirical MSE's come with as. 95%–confidence intervals, which are based on the CLT for the variables

$$\overline{\rm empMSE}_n = \tfrac{n}{10000} \sum\nolimits_j [S_n({\rm sample}_j)]^2 \tag{5.2}$$

Note that with respect to (3.23)/(3.24), and the considered estimators, a contamination point 100 will largely suffice to attain the maximal MSE on $\tilde{\mathcal{Q}}_n$.

5.2 Numerical evaluations

By means of relations (1.14) we may reduce the problem of finding the exact distribution of our M-estimators to the calculation of the "exact" distribution of $\sum_i \psi(X_i)$. For this purpose, we may apply the general convolution algorithm for arbitrarily distributed real-valued random variables introduced in Kohl et al. (2004). This algorithm is based on FFT resp. discrete Fourier Transformation (DFT) and is implemented in `R` within the package `distr` available on `CRAN`, confer Ruckdeschel et al. (2004).
In Ruckdeschel and Kohl (2004), to increase accuracy for M-estimators to Hampel IC's, we extend our algorithm from `distr` to (a) better cope with mass points in $\pm b$ and (b) to calculate the "exact" finite-sample maximum MSE on $\tilde{\mathcal{Q}}_n$. Here we confine ourselves to attach extra columns "numeric" to the following tables summarizing our simulation. "numeric" will then stand for application of Algorithm C respectively Algorithm D from Ruckdeschel and Kohl (2004).
More specifically, for "exact" terms, as worked out in Algorithm C (ibid.), we have to take into account that after conditioning w.r.t. the event that the number of contaminations K in the sample is less than half the sample size, the switching variables U_i from (1.6) no longer are independent. So we may only apply the FFT-based Algorithm from Kohl et al. (2004) to an absolutely continuous inner part and have to calculate the rest by explicitly summing up the events —for details confer Ruckdeschel and Kohl (2004) and the `R`-program written for this purpose, which may be downloaded on the web-page to this article.

On the other side, as described in Algorithm D in Ruckdeschel and Kohl (2004), by the exponential negligibility shown in subsection 2.3, the dependency of the U_i

may be ignored for n sufficiently large —in our case this was possible for $n \geq 30$, moderate radius r and robust clipping height c. Then, we simply may determine the corresponding convolutions of the corresponding distributions of the summands directly by Algorithm 4.4 from Kohl et al. (2004).
To demonstrate the negligibility, for $n \leq 30$, we calculate both "exact" terms (Algortihm C) and those obtained by superposition of the a.c. part and the random walk, ignoring all mass points of the law of the sum (Algortihm D).

5.3 Results

A more detailed account of the results of the simulation study in tables may be found at the web-page to this article. Here we only present some few results which led to the subsequent investigation.

5.3.1 Fixed procedure, fixed radius

To get an idea of the speed of the convergence of the MSE to its as. values, we consider the `H07`-estimator from Andrews et al. (1972), i.e. the M-estimator to $\eta_{0.7}$ at $r = 0.1$ and at $r = 0.5$ for different sample sizes n.
The simulated empirical risk comes with an (empirical) 95% confidence interval and is compared to the corresponding numerical approximations and to the f-o, s-o, and t-o asymptotics from Theorem 3.6. Corresponding tables for the f-o-o M-estimator to η_{c_0} may be drawn from the web-page to this article. The results are tabulated in Tables 2/3. In Table 4 we consider the relative MSE, calculated as the quotient

Table 2: emp., num., and as. MSE at $r = 0.1$, $c = 0.7$

$n/$		simulation			numeric		asymptotics		
situation		$\bar{S}_n$	[low;	up]	Algo C	Algo D	n^0	$n^{-1/2}$	n^{-1}
5	id	1.147	[1.114	;1.179]	1.172	1.168	1.187	1.187	1.169
	cont	1.403	[1.359	;1.447]	1.434	1.535	1.205	1.342	1.345
10	id	1.179	[1.139	;1.205]	1.177	1.174	1.187	1.187	1.178
	cont	1.331	[1.292	;1.369]	1.327	1.326	1.205	1.302	1.303
30	id	1.209	[1.175	;1.242]	1.183	1.180	1.187	1.187	1.184
	cont	1.301	[1.264	;1.337]	1.265	1.262	1.205	1.261	1.261
50	id	1.192	[1.158	;1.225]	–	1.181	1.187	1.187	1.185
	cont	1.250	[1.214	;1.285]	–	1.247	1.205	1.248	1.249
100	id	1.161	[1.128	;1.193]	–	1.182	1.187	1.187	1.186
	cont	1.212	[1.178	;1.246]	–	1.232	1.205	1.236	1.236

$\mathrm{MSE}(c, r)/\mathrm{MSE}(c_0(r), r)$. This is a natural expression to compare the efficiency of different procedures. We compare the empirical terms from the simulation to the corresponding numerical approximations and to the as. terms derived by means of Theorem 3.6. We already recognize a very good approximation down to very small sample sizes.

Table 3: emp., num., and as. MSE at $r = 0.5$, $c = 0.7$

$n/$ situation		simulation $\bar{S}_n$	[low; up]	numeric Algo C	Algo D	asymptotics n^0	$n^{-1/2}$	n^{-1}
5	id	1.166	[1.134 ;1.199]	1.172	1.168	1.187	1.187	1.169
	cont	2.989	[2.892 ;3.087]	3.016	12.491	1.647	2.529	3.103
10	id	1.191	[1.157 ;1.224]	1.177	1.174	1.187	1.187	1.178
	cont	2.934	[2.836 ;3.032]	2.840	4.820	1.647	2.271	2.557
30	id	1.194	[1.161 ;1.227]	1.183	1.180	1.187	1.187	1.184
	cont	2.183	[2.119 ;2.247]	2.167	2.167	1.647	2.007	2.102
50	id	1.165	[1.133 ;1.197]	–	1.181	1.187	1.187	1.185
	cont	1.946	[1.893 ;1.998]	–	2.008	1.647	1.926	1.983
100	id	1.192	[1.159 ;1.226]	–	1.182	1.187	1.187	1.186
	cont	1.894	[1.844 ;1.944]	–	1.879	1.647	1.844	1.873

5.3.2 Fixed procedure, fixed sample size

In order to study the effect of the radius on the quality of the approximation, we consider the M-estimator to $\eta_{0.5}$ at sample size $n = 30$ at varying radii. The results are tabulated in Table 5. The simulations and the numeric values clearly show that with increasing radius, the approximation quality of f-o asymptotics decreases, which is conformal to the infinitesimal character of our neighborhoods. A corresponding table for the more liberal M-estimator to η_2 at sample size $n = 50$ may be drawn from the web-page.

5.3.3 Fixed radius, fixed sample size

In this paragraph we want to compare M-estimators to different clipping heights and see whether the choice of c_0 may also be considered reasonable for moderate n. To this end, we consider the situation $r = 0.25$ and $n = 30$. The results are tabulated in Tables 6 and 7. The simulations already indicate that the answer should be affirmative. The numeric and as. values for the median are taken from Ruckdeschel (2005a). Corresponding tables to the situation $r = 0.5$ and $n = 100$ are on the web-page.

5.3.4 Relative error compared to numerically exact risk

A closer look onto the relative error of our higher order asymptotics w.r.t. the numerically exact risk MSE_n is provided by figure 1. A zoom-in for $n \geq 16$ is available on the web-page. Indeed for all investigated radii $r = 0.00$, 0.10, 0.25, 1.00, the relative error of our asymptotic formula w.r.t. the corresponding numeric figures is quickly decreasing in absolute value in n; also, we notice that we have

Table 4: emp., num., and as. relMSE at $r = 0.1, 0.5$, $c = 0.7$ relative to $\mathrm{Var}[\bar{X}_n]$ for id and $\mathrm{MSE}(c_0(r))$ for cont

n/ situation		$r=0.1$ sim	num ex/∗	asymptotics n^0	$n^{-1/2}$	$r=0.5$ sim	num ex/∗	asymptotics n^0	$n^{-1/2}$
5	id	1.161	1.163	1.173	1.173	1.038	1.042	1.041	1.041
	cont	1.003	0.956	1.143	1.039	0.992	0.978	1.006	0.989
10	id	1.167	1.166	1.173	1.173	1.037	1.041	1.041	1.041
	cont	1.049	1.029	1.143	1.065	0.993	0.977	1.006	0.992
30	id	1.174	1.170	1.173	1.173	1.037	1.041	1.041	1.041
	cont	1.094	1.086	1.143	1.095	0.994	0.993	1.006	0.997
50	id	1.160	1.169*	1.173	1.173	1.038	1.041*	1.041	1.041
	cont	1.096	1.096*	1.143	1.105	0.996	0.995*	1.006	0.999
100	id	1.180	1.170*	1.173	1.173	1.044	1.041*	1.041	1.041
	cont	1.122	1.110*	1.143	1.116	0.999	0.999*	1.006	1.001

Table 5: emp., num., and as. MSE at $n = 30$, $c = 0.5$

r	simulation $\bar{S}_n$	[low; up]	numeric Algo C	Algo D	asymptotics n^0	$n^{-1/2}$	n^{-1}
0.00	1.272	[1.237 ;1.307]	1.259	1.256	1.263	1.263	1.259
0.10	1.374	[1.336 ;1.413]	1.337	1.335	1.280	1.334	1.334
0.25	1.545	[1.502 ;1.588]	1.545	1.542	1.588	1.514	1.532
0.50	2.204	[2.139 ;2.268]	2.189	2.187	1.689	2.037	2.128
1.00	5.362	[5.219 ;5.505]	5.238	5.265	2.967	4.132	4.652

a certain oscillation between odd and even sample sizes for very small n which is explained by the fact that for even n there may be ties. By Lemma 1.1, the contribution of these ties to the risk is however decaying exponentially in n.
In table 8, we have determined the smallest sample size n_0 such that for $n \geq n_0$ the relative error using first to third order asymptotics for approximating $\mathrm{MSE}_n(\psi_c)$ to $c = 0.7$ is smaller than 1% resp. 5% which shows that for $r \leq 0.5$ we need no more than 25 (60) observations to stay within an error corridor of 5% (1%) in t-o asymptotics. For f-o asymptotics, however we need considerable sample sizes for reasonable approximations unless the radius is rather small.
The figures in this table are to be taken "cum grano salis" due to numerical inaccuracies in MSE_n w.r.t. the exact risk of order $1\mathrm{E}-5$ which may result in a deviation from the "real" n_0 of ± 2 for $n_0 < 200$.

Table 6: emp., num., and as. MSE at $n = 30$, $r = 0.25$

estimator/ situation		simulation $\bar{S}_n$	[low; up]	num ex	asymptotics n^0	$n^{-1/2}$	n^{-1}
Med	id	1.492	[1.451 ;1.532]	1.501	1.571	1.571	1.496
	cont	1.786	[1.736 ;1.835]	1.779	1.669	1.821	1.767
$c = 0.5$	id	1.250	[1.216 ;1.284]	1.259	1.263	1.263	1.259
	cont	1.545	[1.502 ;1.588]	1.545	1.369	1.514	1.532
$c = 1.0$	id	1.092	[1.062 ;1.122]	1.105	1.107	1.107	1.105
	cont	1.433	[1.393 ;1.473]	1.440	1.241	1.402	1.425
$c = 2.0$	id	0.991	[0.963 ;1.018]	1.010	1.010	1.010	1.010
	cont	1.611	[1.566 ;1.656]	1.633	1.285	1.556	1.604
$c = c_0 = 1.3393$	id	1.035	[1.006 ;1.063]	1.051	1.139	1.053	1.052
	cont	1.438	[1.398 ;1.479]	1.452	1.220	1.405	1.434

Table 7: emp., num., and as. relMSE at $n = 30$, $r = 0.25$ relative to $\mathrm{Var}[\bar{X}_n]$ for id and $\mathrm{MSE}(c_0(r))$ for cont, $c_0(r) = 1.3393$

estimator/ situation		simulation	numeric ex	asymptotics n^0	$n^{-1/2}$
Med	id	1.435	1.427	1.379	1.379
	cont	1.241	1.224	1.320	1.263
$c = 0.5$	id	1.202	1.197	1.199	1.198
	cont	1.073	1.064	1.077	1.068
$c = 1.0$	id	1.051	1.051	1.051	1.051
	cont	0.995	0.991	0.998	0.994
$c = 2.0$	id	0.953	0.960	0.959	0.960
	cont	1.119	1.125	1.107	1.119

6 Ramifications

6.1 Ideal distributions with polynomially decaying tails

In order to be able to cover ideal distributions with polynomially decaying tails, we sharpen the restriction of the original neighborhood system $\tilde{\mathcal{Q}}_n(r, \varepsilon_0)$ from (2.5) to

$$Q_n = \mathcal{L}\Big\{[(1-U_i)X_i^{\mathrm{id}} + U_i X_i^{\mathrm{di}}]_i \;\Big|\; \limsup_n \tfrac{1}{n}\sum_{i=1}^n U_i \le \varepsilon_0' \Big\} \tag{6.1}$$

for some fixed ε_0' such that

$$0 \le \varepsilon_0' < \varepsilon_0 \tag{6.2}$$

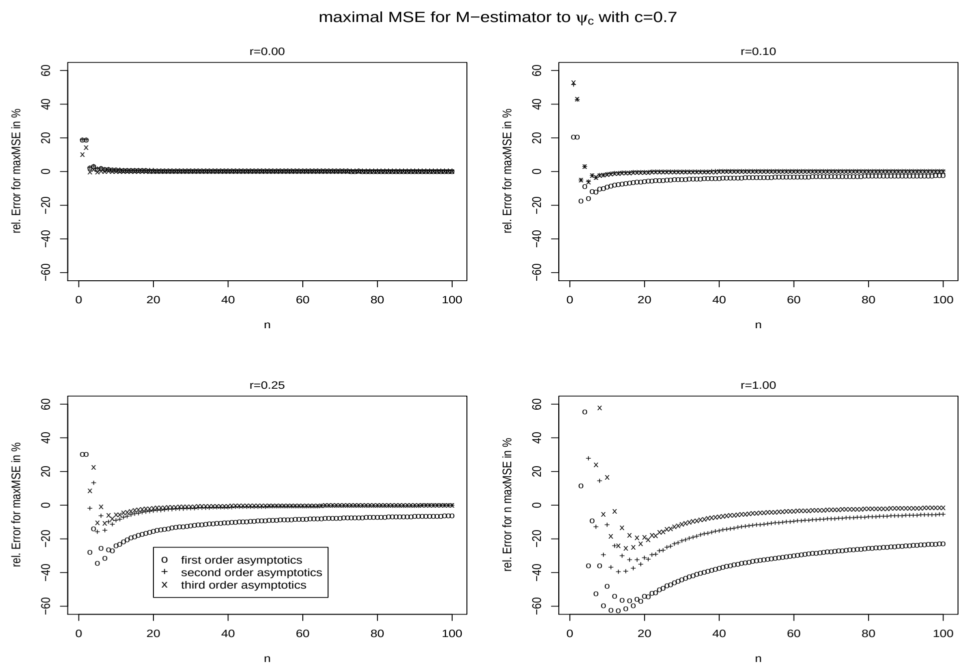

Figure 1: The mapping $n \mapsto \mathrm{rel.error}(\mathrm{MSE}_n(\psi_c))$ for $c = 0.7$ and $F = \mathcal{N}(0,1)$.

Table 8: Minimal n_0 such that for $n \geq n_0$ the relative error using first to third order asymptotics for approximating $\mathrm{MSE}_n(\psi_c)$ for $c = 0.7$ is smaller than 1% resp. 5%

rel.err	order	$r = 0.00$	$r = 0.10$	$r = 0.25$	$r = 0.50$	$r = 1.00$
1%	1st order asy.	9	$> 640^*$	$> 3927^*$	$> 14425^*$	$> 49220^*$
	2nd order asy.	9	15	60	196	$> 580^*$
	3rd order asy.	5	15	30	59	146
5%	1st order asy.	3	28	162	$> 590^*$	$> 1995^*$
	2nd order asy.	3	6	17	43	119
	3rd order asy.	3	6	12	23	49

$*$: for $n > 200$ computation of MSE_n gets too expensive in time; instead we use the the corresponding t-o figure. Assuming an error of t-o asymptotics of order $\mathrm{O}(n^{-3/2})$, a corresponding regression onto the error term gives estimates for the regression coefficient to the term $n^{-3/2}$ of about -50, -166, -534, and -1940 for $r = 0.1$, 0.25, 0.5, and 1.0, so that the error (read from top to bottom and then left to right) incurred by this replacement is about $-3\mathrm{E}{-3}$, $-7\mathrm{E}{-4}$, $-3\mathrm{E}-4$, $-2\mathrm{E}-2$, $-2\mathrm{E}-2$, $-1.3\mathrm{E}-1$, and $-2\mathrm{E}-4$.

giving the new neighborhood system $\tilde{\mathcal{Q}}_n'(r;\varepsilon_0')$. Correspondingly, we will consider the asymptotics of

$$R_n'(S_n, r; \varepsilon_0') := \sup_{Q_n \in \tilde{\mathcal{Q}}_n'(r;\varepsilon_0')} n \int |S_n - \theta_0|^2 \, dQ_n \tag{6.3}$$

It is not surprising that all results up to this point on maximal risks are unaffected by this subtle modification. But, we may replace assumption (Vb) by

(Pd) There are some $T > 0$ and $\eta > 0$ such that

$$F(t) \geq 1 - t^{-\eta}, \quad \text{for } t > T, \qquad F(t) \leq (-t)^{-\eta} \quad \text{for } t < -T \tag{6.4}$$

Proposition 6.1 *In the location model* (1.1) *with* (1.2), *assume (bmi), (D), and (C) from section 3.2; additionally assume that the central distribution F satisfies* (6.4). *Then, on $\tilde{\mathcal{Q}}_n'(r;\varepsilon_0')$, the assertions of Theorem 3.6 —with any $k_2 > 2$— continue to hold.*

Property (6.4) can be made plausible by the following proposition:

Proposition 6.2 *In the location model* (1.1) *with* (1.2), *assume: For any $d > 0$,*

$$\liminf_{t\to\infty} t^d (1 - F(t)) > 0 \qquad \textit{or} \qquad \liminf_{t\to\infty} t^d F(-t) > 0 \tag{6.5}$$

Then for any sample size n, the MSE of the M-estimator S_n to any IC ψ according to (bmi) in the ideal model is infinite.

Conditions (3.23) resp. (3.24) almost characterize the risk-maximizing contaminations:

Proposition 6.3 *Under the assumptions of Theorem 3.6, let $\delta_0, c_0 > 0$. Assume that $\hat{b} = b$ and let $B_n := \inf\{x \,|\, \psi(x) \geq b - c_0/\sqrt{n}\}$. Assume that, for $K = \sum_{i=1}^n U_i$ and $k > (1-\delta)r\sqrt{n}$,*

$$\Pr\Big(\sum_{i=1}^n U_i \, \mathrm{I}(X_i^{\mathrm{di}} \leq B_n + v_0\sqrt{\log(n)/n}) \geq 1 \,\Big|\, K = k\Big) \geq p_0 > 0 \tag{6.6}$$

Then, eventually in n, for any such sequence of contaminations $Q_n^\flat \in \tilde{\mathcal{Q}}(r)$, the maximal MSE as in condition (3.24) *(i.e. with positive bias) in* (3.20) *cannot be attained. More precisely,*

$$R_n(S_n, r) - n \,\mathrm{E}_{Q_n^\flat} S_n^2 \geq 2p_0 v_0 (rc_0 + b)/(n\sqrt{2\pi}) \tag{6.7}$$

A corresponding relation holds for condition (3.23).

6.2 Convergence of variance and bias separately

The technique used to derive Theorem 3.6 also applies if we are interested in variance and bias separately; we get

Proposition 6.4 *Under Assumptions (bmi) to (C) and for sample size n, an M-estimator S_n for scores-function ψ under a measure $Q_n^0 \in \tilde{\mathcal{Q}}_n(r;\varepsilon_0)$ according to* (3.23) *resp.* (3.24) *admits the following expansions*

$$\begin{aligned}
\sqrt{n}\,\Big|\mathrm{Bias}(S_n,Q_n^0)\Big| &= \Big|rb+\tfrac{1}{\sqrt{n}}\,B_{1,0}+\tfrac{r^2}{\sqrt{n}}\,B_{1,1}+\tfrac{r}{n}\,B_2\Big|+\mathrm{o}(n^{-1}) && (6.8)\\
n\,\mathrm{Bias}^2(S_n,Q_n^0) &= r^2b^2+\tfrac{r}{\sqrt{n}}\,C_1+\tfrac{1}{n}\,C_2+\mathrm{o}(n^{-1}) && (6.9)\\
n\,\mathrm{Var}(S_n,Q_n^0) &= v_0^2+\tfrac{r}{\sqrt{n}}\,D_1+\tfrac{1}{n}\,D_2+\mathrm{o}(n^{-1}) && (6.10)
\end{aligned}$$

with

$$\begin{aligned}
B_{1,0} &= (\tfrac{1}{2}l_2+\tilde{v}_1)v_0^2, \qquad B_{1,1} \;=\; b(1\pm\tfrac{1}{2}l_2b) && (6.11)\\
B_2 &= \Big[(\tfrac{1}{2}{l_2}^2+\tfrac{1}{6}l_3)b^3+b\pm l_2b^2\Big]r^2+b(1\pm\tfrac{1}{2}l_2b)+\\
&\quad+\Big[(\tfrac{1}{2}l_3+\tfrac{3}{2}\,{l_2}^2+\tilde{v}_2+\tilde{v}_1^2+3\,\tilde{v}_1\,l_2)b\pm\tfrac{1}{2}l_2\pm\tilde{v}_1\Big]{v_0}^2 && (6.12)\\
C_1 &= b^2r^2(\pm l_2b+2)\pm b(l_2+2\tilde{v}_1)v_0^2 && (6.13)\\
C_2 &= (\tilde{v}_1\,l_2+\tfrac{1}{4}\,{l_2}^2+\tilde{v}_1^2){v_0}^4+\Big[3\,b^2\pm3\,l_2\,b^3+(\tfrac{5}{4}\,{l_2}^2+\tfrac{1}{3}\,l_3)b^4\Big]r^4+\\
&\quad+(\tfrac{7}{2}\,{l_2}^2+l_3+2\,\tilde{v}_2+2\,\tilde{v}_1^2+7\,\tilde{v}_1\,l_2)b^2\,{v_0}^2\,r^2+\\
&\quad\Big[\pm(2\,l_2+4\,\tilde{v}_1)\,b\,v_0^2+2b^2\pm l_2\,b^3\Big]r^2 && (6.14)\\
D_1 &= \Big[\pm2(\,l_2+\,\tilde{v}_1)b+1\Big]{v_0}^2+b^2 && (6.15)\\
D_2 &= (l_3+\tfrac{7}{2}\,{l_2}^2+11\,\tilde{v}_1\,l_2+8\,\tilde{v}_1^2+3\,\tilde{v}_2){v_0}^4+\\
&\quad\Big[\Big((l_3+\tilde{v}_1^2+\tilde{v}_2+5\,\tilde{v}_1\,l_2+4\,{l_2}^2)b^2\pm4(l_2+\tilde{v}_1)b+1\Big){v_0}^2+\\
&\quad\pm2\,l_2\,b^3+3b^2\Big]r^2+\Big(\tfrac{2}{3}\,\rho_1+(l_2+2\,\tilde{v}_1)\rho_0\Big){v_0}^3 && (6.16)
\end{aligned}$$

where we are in the $-\,[+]$ *-case according to whether* (3.23) *or* (3.24) *applies.*

For a proof to this proposition, we may proceed exactly as in the proof of Theorem 3.6; only in (8.57), we keep the integration domain and replace the integrand $u_1(s)^2\,\varphi(s)\,g_n(s)$ by $u_1(s)\,\varphi(s)\,g_n(s)$; we do not spell this out here. In `MAPLE` the expressions are obtained by means of our procedure `asESi`.

6.3 Other loss functions

Analogously, we obtain that under similar condition as for Theorem 3.6, we may replace the integrand $u_1(s)^2\,\varphi(s)\,g(s)$ in (8.57) — on essentially the same domain of integration — using some other loss function ℓ, i.e. by $\ell(u_1)\,\varphi(s)\,g(s)$. In this

respect, Theorem 3.6 easily extends to uniform convergence of other risks on $\tilde{\mathcal{Q}}_n$, e.g. absolute error ($\ell(x)=|x|$), L_k-error ($\ell(x)=|x|^k$) for $1<k<\infty$, and certain covering probabilities, $\ell(x)=\mathrm{I}_{(\alpha_1,\alpha_2)}(x)$ for some $\alpha_1<\alpha_2\in\mathbb{R}$.
As an illustration we consider this last type of loss function, more specifically in the form in which it arises in the finite minimax estimation theory as in Huber (1968) and in which it has been extended to an as. setup by Rieder (1980): The risk is defined as

$$R^\natural(S_n,r)=\sup_{Q_n\in\mathcal{Q}_n(r)}\max\{Q_n(S_n>\theta+\frac{\alpha_2}{\sqrt{n}}),\,Q_n(S_n<\theta-\frac{\alpha_1}{\sqrt{n}})\} \tag{6.17}$$

Recently Fraiman et al. (2001) have taken up a similar setup with conventional confidence intervals to cover bias and variance simultaneously.
We work in the setup of Rieder (1980) here and confine ourselves to the higher order terms of order $n^{-1/2}$, but of course an extension to terms up to order n^{-1} as in Theorem 3.6 is feasible. Due to translation equivariance, it is no restriction to consider the case $\theta=0$ only. As in Rieder (1980), we work with a possibly asymmetric partition of the interval of given length $2a/\sqrt{n}$ laid around the estimator: Using the partition

$$2a=\alpha_1+\alpha_2=\alpha_1(S_n)+\alpha_2(S_n), \tag{6.18}$$

we minimize the risk according to Rieder (1980, formulas (2.8) and (2.11) in), if with $\check{b}$, $\hat{b}$, and $\bar{b}$ from (2.3) and

$$\alpha_1=a-\delta,\qquad \alpha_2=a+\delta,\qquad \delta=\tfrac{r}{2}\,(\hat{b}+\check{b}) \tag{6.19}$$

If we now account for terms of order $\frac{1}{\sqrt{n}}$ we minimize the risk if we use the partition

$$2a=\alpha_1'+\alpha_2'=\alpha_1'(S_n)+\alpha_2'(S_n), \tag{6.20}$$

with

$$\alpha_1'=a-\delta-\delta',\qquad \alpha_2=a+\delta+\delta', \tag{6.21}$$

$\delta'=\delta_n'$ given in the theorem below. To this end, let

$$s_1:=(-a+r\bar{b})/v_0 \tag{6.22}$$

Then, with Φ and φ c.d.f. and density of $\mathcal{N}(0,1)$ and using the notation of Theorem 3.6, we have

Theorem 6.5 *For the location model* (1.1) *of finite Fisher information* (1.2)*, assume (bmi), (D') and (C'). Then for sample size* n*, the minimal over-/undershooting probability of an M-estimator* S_n *for scores-function* ψ *in* $\mathcal{Q}_n$ *obtains eventually in* n *as*

$$\begin{aligned}R^\natural(S_n)=&\sup_{Q_n\in\mathcal{Q}_n}\max\{Q_n(S_n\le-\frac{\alpha_1'}{\sqrt{n}}),\,Q_n(S_n\ge\frac{\alpha_2'}{\sqrt{n}})\}=\\ &=R_-(S_n,Q^0_{n;-})=R_+(S_n,Q^0_{n;+})\end{aligned} \tag{6.23}$$

with $Q^0_{n;-}$ *resp.* $Q^0_{n;+}$ *according to* (3.23) *resp.* (3.24) *and*

$$R_-(S_n, Q^0_{n;-}) = \Phi(s_1) + \tfrac{1}{\sqrt{n}\, v_0}\varphi(s_1)\times \\ \times\Big[\tfrac{ra}{2} + 2l_2 a\delta - a s_1 \tilde v_1 v_0 - \tfrac{r(\check b^2+\hat b^2)s_1}{4v_0} + \tfrac{r^2\bar b}{2}\Big] + \mathrm{o}(\tfrac{1}{\sqrt{n}}) \tag{6.24}$$

and $\delta' = \delta'_n$ *according to*

$$\delta' = \tfrac{1}{\sqrt{n}}\Big(-\tfrac{r\delta}{2v_0} - \tfrac{l_2}{2v_0}(a^2+\delta^2) - \tilde v_1 v_0 s_1\delta - \tfrac{\rho_0}{6}(s_1^2-1) + \tfrac{r\bar b\delta s_1}{v_0^2} + \tfrac{r^2\delta}{2v_0}\Big) \tag{6.25}$$

Remark 6.6 (a) If $l_2 = \tilde v_1 = 0$ and $\hat b = -\check b$, we obtain the same result as (6.24), if we use the expressions $b_n := \mathrm{Bias}_n$ and $v_n^2 = \mathrm{Var}_n$ for bias and variance from Proposition 6.4, plug them into the as. risk, which gives $\Phi((rb_n - a)/v_n)$, and then expand this up to $\mathrm{o}(n^{-1/2})$.

(b) The numerical values obtainable by Theorem 6.5 should be compared to those of Kohl (2005, sections 11.3.3.3 and 11.4.1); admittedly the approach of Theorem 6.5 in this context gives rather poor (too liberal) approximations compared to those in the cited reference (confer the R-file `Thm65.R` available on the web-page to this article).

6.4 Different models

By the Log-Transformation — c.f. Kohl (2005, p. 156-159) —we may transform any scale model to a location model and thus also cover this model directly. For general parametric models arguments like in Rieder (1994, section 6.2.2) seem necessary, and with these in principle corresponding higher order statements should be possible.

7 Consequences

In this section, we consider the class $\mathcal{S}_2$ of all M-estimators according to (bmi), (D'), and (C') as well as (Vb) or (Pd); correspondingly, we define $\mathcal{S}_3$ with (D), (C) replacing (D'), (C'); we always assume that the class of M-estimators $\mathcal{H}$ of ICs of Hampel-type (1.10) forms a subset of $\mathcal{S}_2$ $[\mathcal{S}_3]$.

7.1 Second-order optimality

Symmetry allows considerable simplifications; for instance, if F is symmetric, i.e. $F(B) = F(-B)$ for all $B \in \mathbb{B}$, in (1.10) always $z = 0$. But also, much deeper results are possible. Thus for the rest of this subsection, we assume

$$l_2 = v_1 = \rho_0 = 0 \tag{7.1}$$

Under these assumptions, we come up with (8.76) as s-o-maximal MSE for any M-estimator in $\mathcal{S}_2$; in particular

$$A_1 = v_0^2 + b^2(1 + 2r^2) \tag{7.2}$$

Condition (7.1) is clearly the case for skew symmetric ψ and symmetric F. For symmetric F, however, for any IC ψ, also $\tilde{\psi} := -\psi(-\,\cdot\,)$ is an IC and hence so is the skew-symmetrized $\psi^{(s)} := \frac{1}{2}(\psi + \tilde{\psi})$, too. But by convexity of the MSE, $\psi^{(s)}$ will be at least as good as ψ as to MSE, hence it is no restriction to only consider skew symmetric ICs, and we fall into the application range of

Theorem 7.1 *Assume that maximal as. risk of an ALE on $\tilde{\mathcal{Q}}_n$ resp. $\tilde{\mathcal{Q}}_n'(\,\dot{}\,, s_0)$ is representable as $G(rb(\psi), v_0(\psi))$ for some convex real-valued function $G(w,s)$, strictly isotone in both arguments and totally differentiable, bounded away from the minimum for $w \to \infty$. Then, on $\mathcal{Q}_n$, respectively on $\tilde{\mathcal{Q}}_n$, the optimal IC of Hampel-type* (1.10) *for some clipping height $b = Ac$ determined by*

$$r\, v_0\, \partial_w G(rAc, v_0) = \partial_s G(rAc, v_0)\, A\, \mathrm{E}(|\Lambda - z| - c)_+ \tag{7.3}$$

proved in Ruckdeschel and Rieder (2004). In our case, this theorem specializes to

Corollary 7.2 *Assume a symmetric model* (1.1) *with increasing Λ_f and* (1.2). *Under the assumptions of this section, the s-o-o M-estimator in class $\mathcal{S}_2$ has an IC of of Hampel-type* (1.10) *with $z = 0$ and the s-o-o clipping height $c_1 = c_1(n)$ is determined by*

$$r^2 c\Big(1 + \frac{r^2+1}{r^2 + r\sqrt{n}}\Big) = \mathrm{E}(|\Lambda| - c)_+ \tag{7.4}$$

Always, $c_0 > c_1(n)$. Suppose that $h(c) := \mathrm{E}(|\Lambda| - c)_+$ is differentiable in c_0 with derivative $h'(c_0)$. Then,

$$c_1(n) = c_0\Big(1 - \frac{1}{\sqrt{n}}\,\frac{r^3 + r}{r^2 - h'(c_0)}\Big) + \mathrm{o}(\frac{1}{\sqrt{n}}) \tag{7.5}$$

That is, **the f-o-o clipping height c_0 always is too optimistic.**

Assume s-o risk of ICs of Hampel-type (1.10) is smooth enough in c in its minimum c_1 to allow a s-o Taylor expansion. Then, around c_1, s-o risk behaves like a parabola. But, as by (7.5), $c_1 - c_0 = \mathrm{O}(1/\sqrt{n}\,)$, s-o risk improvement by using c_1 instead of c_0 can only be of order $\mathrm{O}(1/n)$. This even carries over to risks "near" s-o risk:

7.2 Consequences for the exact MSE

Proposition 7.3 *Let $F, F_n, G_n \in \mathcal{C}_2(\mathbb{R})$, $n \in \mathbb{N}$, such that for some $\beta \geq \beta' > 0$*

$$\begin{aligned} &\text{(i)} \quad \sup_x |F_n - G_n| + |F_n' - G_n'| + |F_n'' - G_n''| = \mathrm{O}(n^{-\beta}),\\ &\text{(ii)} \quad \sup_x |F_n - F| + |F_n' - F'| + |F_n'' - F''| = \mathrm{O}(n^{-\beta'}) \end{aligned} \tag{7.6}$$

Assume that in $x_0 \in \mathbb{R}$, $F(x_0)$ is minimal, and that $F''(x_0) = f_2 > 0$. Then

(a) *there is some sequence $(x_n) \subset \mathbb{R}$ such that eventually in n, $F_n(x_n)$ is minimal and $\lim F_n''(x_n) = f_2$.*

(b) $|x_n - x_0| = \mathrm{O}(n^{-\beta'})$.

(c) *there is some sequence* $(y_n) \subset \mathbb{R}$ *such that eventually in* n, $G_n(y_n)$ *is minimal and* $\lim_n G_n''(y_n) = f_2$.
(d) $|y_n - x_n| = \mathrm{O}(n^{-\beta})$.
(e) $0 \le G_n(x_n) - G_n(y_n) = \mathrm{O}(n^{-2\beta})$.

The drawback of this proposition is that assumption (7.6) is difficult to check if we have no explicit expression for G_n: For given $r \ge 0$, let $\mathrm{asMSE}_{j=0,1,2}(c)$ be the f-o, s-o, and t-o maximal MSE of an M-estimator in $\mathcal{H}$, and $\mathrm{exMSE}(c)$ the corresponding exact maximal MSE R_n; we would like to apply Proposition 7.3 to $F = \mathrm{asMSE}_0$, $F_n = \mathrm{asMSE}_{j=1,2}$ and $G_n = \mathrm{exMSE}$ to conclude on the performance of f-o-o, s-o-o, t-o-o procedures as to exMSE. As to (7.6), part (ii) is easy to see checking the expressions, giving $\beta' = 1/2$, while for part (i) Theorem 3.6 only says that $\sup_x |F_n - G_n| = \mathrm{o}(n^{-j/2})$ which in fact is $\mathrm{O}(n^{-(j/2+\delta)})$, and probably, under slightly stronger assumptions, $\mathrm{O}(n^{-(j+1)/2})$. So presumably —in view of Table 1,

$$0 \le \mathrm{exMSE}(c_{j,n}) - \mathrm{exMSE}(c_{\mathrm{ex};n})) = \mathrm{O}(n^{-j-1}), \qquad j = 0,1,2 \tag{7.7}$$

Remark 7.4 We even conjecture that we may apply an analogue to Proposition 7.3 for functions $F, F_n, G_n\colon \Psi \to \mathbb{R}$: Let us denote by $\hat\psi^{(j;n)}$, the corresponding f-o, s-o, t-o optimal IC and $\hat\psi^{(\mathrm{ex};n)}$ the exactly optimal IC; then, with the usual abuse of notation as to exMSE, we conjecture that

$$0 \le \mathrm{exMSE}(\hat\psi^{(j;n)}) - \mathrm{exMSE}(\hat\psi^{(\mathrm{ex};n)}) = \mathrm{O}(n^{-j-1}), \qquad j = 0,1,2 \tag{7.8}$$

7.3 Relative risk

An observation in the simulation study was that the relative MSE w.r.t. the MSE of the f-o-o procedure seemed to converge faster than the absolute terms. This is reflected by our formulas as follows:

7.3.1 Contaminated situation

Let $\mathrm{asMSE}_0(c)$ and $A_1(c)$ be the f-o as. MSE and the corresponding s-o correction term for the Hampel-IC with clipping height c. Then we may write for the f-o [s-o] relative risk $\mathrm{relMSE}_0(c,r)$ $[\mathrm{relMSE}_1(c,r,n)]$ w.r.t. the corresponding risk of the f-o-o procedure

$$\mathrm{relMSE}_1(c,r,n) := \frac{\mathrm{asMSE}_0(c) + \frac{r}{\sqrt{n}} A_1(c)}{\mathrm{asMSE}_0(c_0) + \frac{r}{\sqrt{n}} A_1(c_0)} = \tag{7.9}$$

$$= \mathrm{relMSE}_0(c,r)\left(1 + \frac{r}{\sqrt{n}}(\Delta(c) - \Delta(c_0))\right) + \mathrm{o}(n^{-1/2}) \tag{7.10}$$

with

$$\Delta(c) := \frac{b^2(c) - v_0^2(c)}{\mathrm{asMSE}_0(c)} \tag{7.11}$$

So in fact, the assumed faster convergence is not true, but as we will see, the difference between $\mathrm{relMSE}_0(c,r)$ and $\mathrm{relMSE}_1(c,r)$ are in fact small.
As we will base our decision which procedure to take upon this relative risk, it is interesting to consider the maximal error w.r.t. the s-o approximation one has to take into account when we use the f-o asymptotics instead. In view of subsection 7.1 we will limit ourselves to only considering Hampel-IC's with a clipping height c in the range

$$C(c_0,\rho) := [c_0/(1+\rho),\, c_0(1+\rho)], \tag{7.12}$$

for $\rho \geq 0$. This leads us to

$$\widehat{\Delta\mathrm{relMSE}}(r;\rho) := \max_{c\in C(c_0(r),\,\rho)} r\,\big(\Delta(c) - \Delta(c_0(r))\big) \tag{7.13}$$

or even maximizing over the radius

$$\widehat{\Delta}(\rho) := \widehat{\widehat{\Delta\mathrm{relMSE}}}(\rho) := \max_r \widehat{\Delta\mathrm{relMSE}}(r;\rho) \tag{7.14}$$

In the Gaussian case, the function $r \mapsto \widehat{\Delta\mathrm{relMSE}}(r;\rho)$ is plotted for $\rho = 0.1$ in Figure 2, and for $\widehat{\Delta}(0.1)$, we get a value of 0.065, which for an actual sample size n has to be divided by $\sqrt{n}$ — an astonishingly good approximation!
So down to very moderate sample sizes we can base our decision which clipping height to take to achieve "nearly" the optimal MSE on $\tilde{\mathcal{Q}}_n$ on f-o asymptotics only.

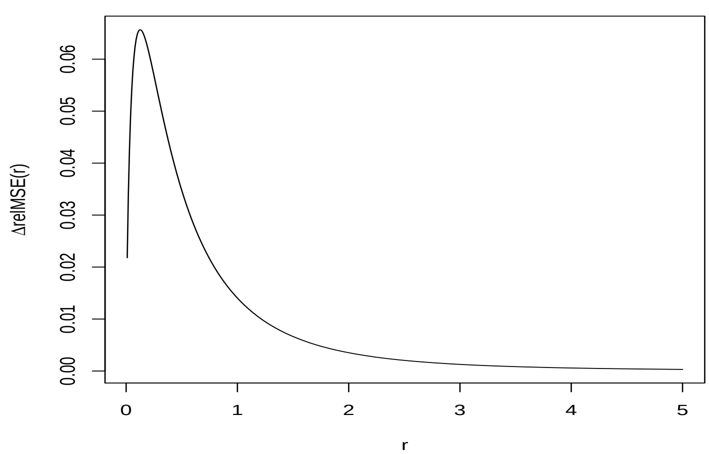

Figure 2: The mapping $r \mapsto \widehat{\Delta\mathrm{relMSE}}(r;\rho)$ for $F = \mathcal{N}(0,1)$ and for $\rho = 0.1$.

7.3.2 Illustration

As an example we take $F = \mathcal{N}(0,1)$ and calculate the terms c_1,

$$\mathrm{asMSE}_1 := \mathrm{asMSE}_0 + \tfrac{r}{\sqrt{n}} A_1 \tag{7.15}$$

and relMSE_1 for the radii and sample sizes of the simulation study where for the optimization for c_1 we use the function `optimize` in `R 1.7.1` (compare R Development Core Team (2005)). The results are tabulated in Table 9. Correspondingly, we also determine the t-o terms c_2,

$$\mathrm{asMSE}_2 := \mathrm{asMSE}_1 + A_2/n \tag{7.16}$$

and in Figure 3, we plot the graphs of the five functions

$$r \mapsto \mathrm{asMSE}_0(\eta_{c_0(r)}, r), \qquad r \mapsto \mathrm{asMSE}_1(\eta_{c_0(r)}, r, n), \qquad r \mapsto \mathrm{asMSE}_2(\eta_{c_0(r)}, r, n)$$
$$r \mapsto \mathrm{asMSE}_1(\eta_{c_1(r,n)}, r, n), \quad r \mapsto \mathrm{asMSE}_2(\eta_{c_2(r,n)}, r, n)$$

for $F = \mathcal{N}(0,1)$ and for $n = 30$. In fact, the choice of the clipping height — $c_0(r), c_1(r,n), c_2(r,n)$ — does not entail any visible changes while the absolute value of f-o, s-o, and t-o MSE clearly differ.
In the same situation, the three functions $r \mapsto c_0(r)$, $r \mapsto c_1(r,n)$, $r \mapsto c_2(r,n)$ are plotted in Figure 4; while there are visible differences between $c_0(r)$ and $c_i(r,n)$, $i = 1,2$, $c_1(r,n)$ and $c_2(r,n)$ visually coincide.

Table 9: $c_1(r,n)$, $\mathrm{asMSE}_1(c_1(r,n),r,n)$ and $\mathrm{relMSE}_1(c_1(r,n),r,n)$

r		$n=5$	$n=10$	$n=30$	$n=50$	$n=100$	$n=\infty$
	c_1	1.394	1.484	1.611	1.663	1.724	1.948
0.1	asMSE_1	1.248	1.197	1.140	1.122	1.103	1.054
	relMSE_1	3.476%	2.149%	0.939%	0.623%	0.349%	0.000%
	c_1	0.994	1.059	1.147	1.181	1.219	1.339
0.25	asMSE_1	1.635	1.519	1.397	1.358	1.319	1.220
	relMSE_1	2.377%	1.470%	0.632%	0.414%	0.228%	0.000%
	c_1	0.650	0.690	0.746	0.767	0.790	0.862
0.5	asMSE_1	2.527	2.271	2.006	1.923	1.840	1.636
	relMSE_1	1.214%	0.772%	0.342%	0.226%	0.126%	0.000%
	c_1	0.320	0.340	0.369	0.380	0.394	0.436
1.0	asMSE_1	5.761	4.944	4.110	3.852	3.593	2.964
	relMSE_1	0.427%	0.292%	0.142%	0.098%	0.056%	0.000%

7.4 Minimax radius

In this subsection, we refine the results of Rieder et al. (2001). In the cited paper, we want to give a guideline to the statistician which procedure to choose if he knows

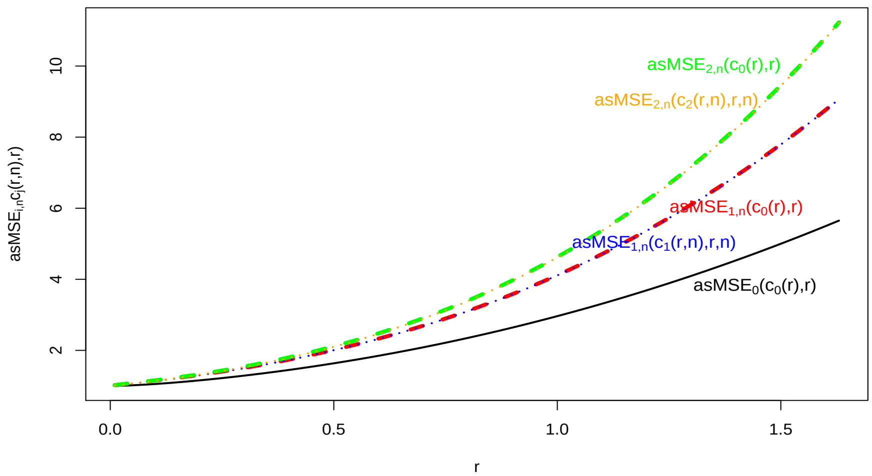

Figure 3: The mapping $r \mapsto \mathrm{asMSE}_{i[,n]}(\eta_{c_j(r[,n])}, r[,n])$ for $i = 0,1,2$, $j = 0, i$, $n = 30$ and $F = \mathcal{N}(0,1)$

that there is contamination but does not know the radius exactly: To this end, we consider the maximal inefficiency $\bar\rho(r')$ defined as

$$\bar\rho_0(r') := \sup_{r\in(r_l,r_u)} \bar\rho(r',r), \qquad \bar\rho(r',r) := \frac{\bar R(\eta_{c_0(r')}, r)}{\bar R(\eta_{c_0(r)}, r)} \tag{7.17}$$

and determine the minimax radius r_0 as minimizer of $\bar\rho_0(r')$. If one knows at least that the actual radius will lie in an interval $[r/\gamma, r\gamma]$ we may determine $r_{\gamma,r}$ as minimizer of $\bar\rho_\gamma(r',r) = \sup_{s\in(r/\gamma,r\gamma)} \bar\rho(r',s)$ and denote the corresponding minimax inefficiency by $\bar\rho_\gamma(r)$. In a second optimizing step we then determine the maximizer r_γ of $\bar\rho_\gamma(r)$. The unrestricted case is symbolically included by $\gamma = \infty$. In the Gaussian location case this gives

$\gamma = 0$			$\gamma = 2$			$\gamma = 3$		
r_0	$c_0(r_0)$	$\bar\rho_0(r_0)$	r_2	$c_0(r_2)$	$\bar\rho_2(r_2)$	r_3	$c_0(r_3)$	$\bar\rho_3(r_3)$
0.621	0.718	18.07%	0.575	0.769	8.84%	0.549	0.799	4.41%

These calculations can easily be translated to the s-o setup setting

$$R_1(\psi, r, n) := r^2 \sup|\psi|^2 + \mathrm{E}\,\psi^2 + \tfrac{r}{\sqrt{n}} A_1 \tag{7.18}$$

so that in this paper we would instead determine $r_1(n)$ as minimizer of $\rho_1(r',r,n)$,

$$\sup_{r\in(r_l,r_u)} \rho_1(r',r,n), \qquad \rho_1(r',r,n) := \frac{R_1(\eta_{c_1(r'(n),n)}, r, n)}{R_1(\eta_{c_1(r,n)}, r, n)} \tag{7.19}$$

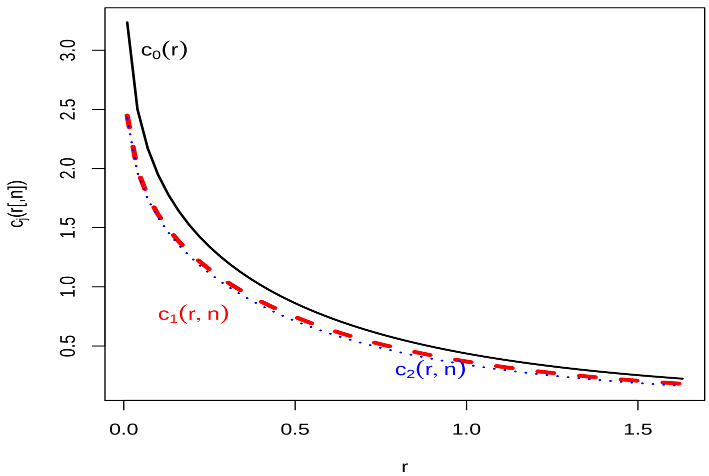

Figure 4: The mapping $r \mapsto c_j(r[,n])$ for $j = 0, 1, 2$ and $n = 30$ and $F = \mathcal{N}(0, 1)$

respectively $\rho_{1;\gamma}$ and instead of $\bar{\rho}_\gamma$. For finite n, however, we have to take into account that $r < \sqrt{n}$ always. Doing so we get Table 10 on page 33, showing that there is not much variation in both $c_1(r_\infty, \cdot)$, $\rho_{1;\gamma}(r_\gamma, \cdot)$ for varying n.
So if r is completely unknown, it is a good choice to use the M-estimator to Hampel-scores for $c \approx 0.7$ —you will never have a larger inefficiency than the limiting 18%! Ex post this is one more argument, why the `H07`-estimate survived in in Sections 7.B.8 and 7.C.4 of the Princeton robustness study (Andrews et al. (1972)). A table for the corresponding t-o minimax radii is available on the web-page.

7.5 Innocent-looking risk-maximizing contaminations

In Huber (1997, p. 62), the author complains "...the considerable confusion between the respective roles of diagnostics and robustness. The purpose of robustness is to safeguard against deviations from the assumptions, in particular against those that are near or below the limits of detectability." As worked out in Ruckdeschel (2004), the exact critical rate for these limits may be determined in a statistical way: For some prescribed outlier set OUT, let p_0 and $q_n = (1 - r_n)p_0 + r_n$ be the probability under the ideal model, and under convex contaminations of radius r_n, respectively. Considering the minimax test between these alternatives yields the exact critical rate $1/\sqrt{n}$: under a faster shrinking p_0 cannot be separated from q_n at all, while at a slower rate, asymptotically we can separate them without error.

Table 10: Minimax radii for second order asymptotics

		$n=5$	$n=10$	$n=30$	$n=50$	$n=100$	$n=\infty$
	r_γ	0.390	0.449	0.514	0.536	0.559	0.621
$\gamma=0$	$c_1(r_\gamma)$	0.776	0.749	0.729	0.725	0.722	0.718
	$\rho_{1;\gamma}(r_\gamma)$	16.27%	17.08%	17.71%	17.85%	17.96%	18.07%
	r_γ	0.481	0.496	0.518	0.524	0.534	0.548
$\gamma=3$	$c_1(r_\gamma)$	0.670	0.694	0.724	0.739	0.750	0.800
	$\rho_{1;\gamma}(r_\gamma)$	6.213%	6.773%	7.490%	7.751%	8.036%	8.836%
	r_γ	0.540	0.552	0.564	0.563	0.571	0.574
$\gamma=2$	$c_1(r_\gamma)$	0.609	0.637	0.675	0.695	0.707	0.770
	$\rho_{1;\gamma}(r_\gamma)$	2.987%	3.297%	3.692%	3.834%	3.988%	4.410%

Going one step further, for some given $1/\sqrt{n}$-shrinking neighborhoods of radius r, we would also like to know how "small" an outlier may be, while it is still harmful enough to distort the classically optimal procedure in a way that this procedure is beaten by some robust one.

7.5.1 The *Cniper* contaminaton

To a fixed radius r, in the preceding sections, we have found/discussed f-o-o and s-o-o ICs of Hampel-form with clipping height $c_j = c_j(r[,n])$, $j=0,1$. To these ICs we have derived families of contaminations achieving maximal risk on $\tilde{\mathcal{Q}}_n(r)$. By means of Theorem 3.6(b), these are induced by any contaminating measures P_n^{di} under which $\eta_\theta(X^{\mathrm{di}})$ is constantly either b_j or $-b_j$ for $b_j = A_j c_j$ —up to an event of probability $\mathrm{o}(n^{-1})$. Out of these risk-maximizing contaminations, let us limit ourselves to those induced by Dirac masses at x:

$$Q_n(x) := [(1-\tfrac{r}{\sqrt{n}})P_\theta + \tfrac{r}{\sqrt{n}}\,\mathrm{I}_{\{x\}}]^{\otimes n} \tag{7.20}$$

Among these $Q_n(x)$, we seek the least "conspicious" looking contamination point x in the sense that the region $\mathrm{OUT}_j := [x;\infty)$ [or $(-\infty;x)$] carries large ideal probability. With this region as outlier set in Ruckdeschel (2004), values of x (or slightly above in absolute value) occuring more frequently than they should under the ideal situation, are hardest to detect.

More precisely, recall the general setup from section 1.3. Assume that the observations are univariate; let $S_n^{(b_0)}$ and $\hat{S}_n$ be ALEs to the classical optimal IC $\hat{\eta} = \mathcal{I}^{-1}\Lambda$ and the asMSE_0-optimal IC η_{b_0}, respectively. In this setup we define

Definition 7.5 *The* f-o cniper point x_0 *is defined as* $x_{0,+}$ *if* $x_{0,+} \ge -x_{0,-}$ *and* $x_{0,-}$ *else, where*

$$\begin{aligned} x_{0,+} &:= \inf\{x>0 \,\big|\, \mathrm{asMSE}_0(S_n^{(b_0)}, Q_n(x)) < \mathrm{asMSE}_0(\hat{S}_n, Q_n(x))\} \\ x_{0,-} &:= \sup\{x<0 \,\big|\, \mathrm{asMSE}_0(S_n^{(b_0)}, Q_n(x)) < \mathrm{asMSE}_0(\hat{S}_n, Q_n(x))\} \end{aligned} \tag{7.21}$$

Remark 7.6 (a) The name *cniper* point is due to H. Rieder; it alludes to the fact that this "Ianus-type" contamination $Q_n(x_0)$ pretends to be *nice*, but to the contrary is in fact *pernic*ious, "sniping" off the classically optimal procedure...

(b) To get rid of the dependency upon the radius r, in the examples we will use the minimax radii $r_\gamma(n)$ defined in the preceding section.

(c) The idea of specifying a contamination appearing as "least dangerous" is of course not bound to quadratic loss.

(d) In the obvious manor, the concept may be generalized for multivariate observations, if we define any x_0 of minimal absolute as *cniper* point.

Correspondingly, in the setup of this paper and under (7.1), let $S_n^{(c_1)}$ be an M-estimator to the s-o-o IC η_{c_1} according to Corollary 7.2.

Definition 7.7 *The* s-o cniper point x_1 *is defined as* $x_{1,+}$ *if* $x_{1,+} \geq -x_{1,-}$ *and* $x_{1,-}$ *else, where*

$$\begin{aligned} x_{1,+} &:= \inf\{x > 0 \,\big|\, \mathrm{asMSE}_1(S_n^{(c_1)}, Q_n(x)) < \mathrm{asMSE}_1(\hat{S}_n, Q_n(x))\} \\ x_{1,-} &:= \sup\{x < 0 \,\big|\, \mathrm{asMSE}_1(S_n^{(c_1)}, Q_n(x)) < \mathrm{asMSE}_1(\hat{S}_n, Q_n(x))\} \end{aligned} \tag{7.22}$$

Cniper contaminations and f/s-o-o ICs form saddle-points under (7.23)/(7.1):

Proposition 7.8 *The pair* $(S_n^{(b_0)}, Q_n(x_0))$ *is a saddlepoint for the class of all pairs* (S_n, Q_n) *if*

$$|\hat{\eta}(x_0)| \leq |\eta_b(x_0)| \qquad \forall b\colon \quad |\eta_b(x_0)| < b \tag{7.23}$$

where S_n *are ALE's to IC's of form* (1.9) *and* $Q_n \in \mathcal{Q}_n$ *w.r.t. f-o risk* $\tilde{R}$.
Under (7.1)*, the same holds in the one-dimensional location model for the pair* $(S_n^{(c_1)}, Q_n(x_1))$ *w.r.t. s-o risk in* $\tilde{\mathcal{Q}}(r)$.

Remark 7.9 A sufficient condition for (7.23) is that $\Lambda(x) = -\Lambda(-x)$: Then for any $b > 0$, $a_b = 0$ is possible and,

$$A_b^{-1} = \mathrm{E}\,\Lambda\Lambda^\tau \min\{1, \frac{b}{|A_b\Lambda|}\} \leq \mathrm{E}\,\Lambda\Lambda^\tau = \mathcal{I}$$

So $A_b \succeq \mathcal{I}^{-1}$ in the positive semi-definit sense, and hence for b s.t. $|\eta_b(x_j)| < b$

$$|\eta_b(x_j)| = |A_b\Lambda(x_j)| \geq |\mathcal{I}^{-1}\Lambda(x_j)| = |\hat{\eta}(x_j)| \tag{7.24}$$

7.5.2 Error probabilities

For numerical evaluations, we consider the Gaussian location model and the Gaussian location and scale model. In both models, $x_{j,+} = -x_{j,-}$, and without loss, we use $x_{j,+}$.
For the as. tests between $q_n = p_0$ and $q_n > p_0$, alluded to in the beginning of this section, we note that

$$p_0 = P_\theta(X_i \geq x_j) = \Phi(-x_j), \qquad q_n = p_0 + \frac{r}{\sqrt{n}}(1 - p_0) \tag{7.25}$$

As to the (f-o) as. minimax test Ruckdeschel (2005b, formula (6.1)) gives as as. risk

$$\varepsilon = \varepsilon_\infty = \Phi\Big(-\frac{r}{2}\sqrt{\frac{1-p_0}{p_0}}\Big) \tag{7.26}$$

For s-o asymptotics, we instead use the finite-sample minimax test, i.e. the Neyman-Pearson test with equal Type-I and Type-II error. In our case this is a corresponding randomized binomial test.

7.5.3 Gaussian location

In the Gaussian location model, we draw all necessary expressions from Proposition 3.4; in particular, with $c_1 = c_1(n, r_\gamma)$, and $A_1 = (2\Phi(c_1)-1)^{-1}$, $b_1 = c_1 A_1$, by Remark 3.7(a), maximizing risk amounts to either $X^{\rm di} > c_1$ always or $X^{\rm di} < -c_1$ always. The classically optimal estimator is the arithmetic mean, and one easily calculates

$$\mathrm{E}_{Q_n(x)}[\bar x_n^2 \,\Big|\, K=k] = \frac{1}{n^2}[k^2x^2 + (n-k)] \tag{7.27}$$

and integrating out K we get directly

$$n\ \mathrm{E}_{Q_n(x)}[\bar x_n^2] = 1 - \tfrac{r}{\sqrt{n}} + x^2(r^2 + \tfrac{r}{\sqrt{n}} - \tfrac{r^2}{n}) \tag{7.28}$$

Combining this with formulas (3.20) and (7.2), for $M_0 := \mathrm{asMSE}_0(S_n^{(c_1)})$ we get

$$x_1^2(n) = \frac{M_0 - 1 + \frac{r}{\sqrt{n}}(M_0 + b_1^2(r^2+1)+1)}{r^2(1-\frac{1}{n}) + \frac{r}{\sqrt{n}}} \tag{7.29}$$

or

$$x_1(n) = \frac{\sqrt{M_0-1}}{r} + \frac{1}{2\sqrt{n}}\Big[\frac{M_0+1+b_1^2(r^2+1)}{\sqrt{M_0-1}} - \frac{\sqrt{M_0-1}}{r^2}\Big] + \mathrm{o}(\tfrac{1}{\sqrt{n}}) \tag{7.30}$$

This yields the results as in Table 11. We include the type-II error $1-\beta(\alpha)$ for the Neyman Pearson test to niveau $\alpha = 5\%$ and the risk ε_n of the corresponding minimax test; roughly speaking we cannot do better than overlooking one of 10 contaminations at niveau 5% ideal observations to be falsely marked as outliers, and, equally weighting the two error types we cannot do better than with a false classification rate of 7% for each error type.

7.5.4 Gaussian location and scale

To give one more example, consider the one-dimensional location-scale model at central distribution $\mathcal{N}(0,1)$. For this model we have not yet established a s-o as. theory; for f-o asymptotics, however, we may use `R`-programs from the bundle `RobASt`, confer Kohl (2005, Appendix D), and get $r_\infty = 0.579$,

$$\max_{Q_n \in \mathcal{Q}_n(r_\infty)} \mathrm{asMSE}(\eta_{\theta;0}, Q_n) = 3.123 \tag{7.31}$$

while $\mathcal{I}_\theta^{-1}\Lambda_\theta = (x, \frac{1}{2}(x^2-1))^\tau$. This gives $x_0 = 1.844$ — and hence $\varepsilon_\infty = 5.737\%$ and $1-\beta_\infty(5\%) = 6.557\%$. Condition (7.23) is proved to hold in subsection 8.11.

Table 11: Minimax contamination at $\gamma = 0$

n	5	10	30	50	100	200	300	∞
$r_\gamma(n)$	0.390	0.449	0.514	0.536	0.559	0.576	0.584	0.621
$c_1(r_\gamma, n)$	0.776	0.749	0.729	0.725	0.722	0.720	0.719	0.718
$x_1(n)$	2.931	2.470	2.101	2.004	1.914	1.853	1.826	1.714
$1 - \beta_n(0.05)$	0.364	0.272	0.215	0.183	0.162	0.133	0.132	0.101
ε_n	0.277	0.178	0.129	0.115	0.097	0.089	0.086	0.072

8 Proofs

8.1 Proof of Proposition 2.1

The assertion (2.2) for uniform normality is Rieder (1994, Theorem 6.2.4). Convergence failure (2.1) is the usual breakdown point argument: W.l.o.g. take $\theta = 0$ .Let $p_n := \Pr(U_i > n/2)$; take x_n so that either $\psi(x_n - \sqrt{K_n/p_n}\,) \uparrow b$ or $\psi(x_n + \sqrt{K_n/p_n}\,) \downarrow -b$. We consider only the first case, here; for the second case, one has to consider $Q_n(S_n \le -t)$. By the relations of Huber (1981, pp. 45), compare (1.14),

$$\begin{aligned} Q_n(S_n \ge t) &\ge Q_n(\textstyle\sum_i \psi(X_i - t) > 0), &(8.1)\\ Q_n(S_n > t) &\le Q_n(\textstyle\sum_i \psi(X_i - t) \ge 0) &(8.2)\end{aligned}$$

Thus for any $t \le \sqrt{K_n/p_n}\,$,

$$\begin{aligned} &Q_n(S_n \ge t) \ge Q_n(\textstyle\sum_i \psi(X_i - t) > 0) \ge \\ &\quad\ge \sum_{k>n/2} \Pr(\textstyle\sum_i \psi(X_i - t) > 0, \sum U_i = k) = \\ &\quad= \sum_{k>n/2} \Pr\Big(\sum U_i = k, \sum_{U_i=0} \psi(X_i - t) > -k\psi(x_n - t) = -kb + \mathrm{o}(n^0)\Big) \quad (8.3)\end{aligned}$$

But, as $\sup|\psi| \le b$ for all $t \in \mathbb{R}$ and all $k > n/2$,

$$\inf_{y_1,\ldots,y_{n-k} \in \mathbb{R}^{n-k}} \sum_{i=1}^{n-k} \psi(y_i - t) \ge -(n-k)b > -kb \tag{8.4}$$

so that for n sufficiently large and for $t \le \sqrt{K_n/p_n}$

$$Q_n(S_n \ge t) \ge \sum_{k>n/2} \Pr(\textstyle\sum U_i = k) = \Pr(U_i > n/2) = p_n \tag{8.5}$$

Now take $t_n := (K_n/p_n)^{1/2}$ to get

$$\mathrm{E}_{Q_n}[S_n^2] \ge t_n^2 Q_n(S_n \ge t_n) \ge K_n \tag{8.6}$$

Here we use the fact, that although arbitrarily small for large n, $p_n > 0$. ////

8.2 Proof to Lemma 3.2

Let G_t be the law of $\psi_t(X^{\mathrm{id}})$. By assumption, the Lebesgue decomposition yields $dG_0 = ag\,d\lambda + (1-a)\,d\tilde G$ for $a \in (0,1]$, g some probability density and $\tilde G \perp \lambda$. The support of g contains an open interval (c_1, c_2) and $G_0(c_2) > G_0(c_1)$. On (c_1, c_2), ψ is strictly isotone and continuous, so that with $d_i = \psi^{-1}(c_i)$

$$P(\psi_t(X^{\mathrm{id}}) \in (c_1,c_2)) = P(d_1 + t < X^{\mathrm{id}} < t + d_2) = \int_{d_1+t}^{d_2+t} dF \tag{8.7}$$

But

$$\int_{d_1+t}^{d_2+t} dF = G_0(c_2) - G_0(c_1) + \mathrm{o}(t^0) \tag{8.8}$$

so that for t small enough, the absolute continuous part of G_t is uniformly bounded away from 0 and hence by the Lebesgue Lemma our condition (3.9) holds. ////

8.3 Proof to Proposition 3.4 and Remark 3.5

To get $\mathrm{E}[\hat\eta_c \Lambda_f] = 1$, the Lagrange multiplier A_c must be determined by

$$A_c^{-1} = 2\Phi(c) - 1$$

It holds that $b = A_c c$. For $c \to \infty$ we obtain the classically optimal IC, and $c \to 0$, using l'Hospital yields the IC of the sample median. As to $L(t)$, we obtain

$$\begin{aligned}
&L_c(t) = A[c - (c+t)\Phi(t+c) + (t-c)\Phi(t-c) + \varphi(t-c) - \varphi(t+c)],\\
&L_\infty(t) = -t, \qquad L_0(t) = \sqrt{\tfrac{\pi}{2}}\,(1 - 2\Phi(t))
\end{aligned}$$

all arbitrarily often differentiable functions, so (3.5) holds with l_i as stated in the proposition. For $V(t)$ introduce

$$S(t) := \mathrm{E}[\psi(x-t)^2], \qquad W(t) := V(t)^2$$

Then, suppressing the argument t,

$$W = S - L^2, \qquad W' = S' - 2LL', \qquad W'' = S'' - 2L'^2 - 2LL''$$

and with $W_0 = W(0)\,,\tilde W_1(0) = W'(0)/W_0\,,\tilde W_2(0) = W''(0)/W_0$ we get

$$W_0 = S(0), \qquad \tilde W_1 = S'(0)/S(0), \qquad \tilde W_2 = (S''(0) - 2)/S(0)$$

and hence

$$V(t) \quad = \quad \sqrt{W_0}\big(1 + \tfrac{\tilde W_1\, t}{2} + \tfrac{(2\tilde W_2 - \tilde W_1^2)\, t^2}{8}\big) + \mathrm{O}(t^{2+\delta})$$

so that

$$v_0 = \sqrt{S(0)}, \qquad \tilde v_1 = \frac{S'(0)}{2S(0)}, \qquad \tilde v_2 = \frac{2S''(0) - 4 - S'(0)^2/S(0)}{4S(0)}$$

In our case we have for $0 < c < \infty$

$$\begin{aligned}
S(t) \quad = \quad & A_c^2\Big[c^2\big(1 - \Phi(t+c) + \Phi(t-c)\big) + (1+t^2)\big(\Phi(t+c) - \Phi(t-c)\big) + \\
& +(t-c)\varphi(t+c) - (t+c)\varphi(t-c)\Big]
\end{aligned}$$

and

$$S(t) = 1 + t^2 \quad \text{for } c = \infty, \qquad S(t) = \frac{\pi}{2} = b^2 \quad \text{for } c = 0$$

so (3.6) holds with

	$0 < c < \infty$	$c = 0$	$c = \infty$
$S(0)$	$2b^2(1-\Phi(c)) + A_c(1-2b\varphi(c))$	1	$\frac{\pi}{2}$
$S'(0)$	0	0	0
$S''(0)$	$2A_c^2(2\Phi(c) - 1 - 2c\varphi(c))$	2	0

and the assertions as to v_0, $\tilde v_1$, $\tilde v_2$ follow. As to (Vb), for $|t| \to \infty$, we get with Mill's ratio for any $\delta > 0$

$$\begin{aligned} \Big| b - |L(t)| \Big| &= A_c \Big| (c+t)\bar\Phi(t+c) - (t-c)\bar\Phi(t-c) + \varphi(t-c) - \varphi(t+c) \Big| = \\ &= \mathrm{o}(\exp(-\frac{t^2}{2+\delta})) \end{aligned}$$

Again with Mill's ratio,

$$|S(t) - b^2| \le A_c^2 \Big[2(t^2+1)\bar\Phi(|t|-c) + 2(|t|+c)\varphi(|t|-c) \Big] = \mathrm{o}(\exp(-\frac{t^2}{2+\delta}))$$

and hence

$$V^2(t) = S(t) - L(t)^2 = \mathrm{o}(\exp(-\frac{t^2}{2+\delta}))$$

For $c = 0$ we get

$$\Big| b - |L(t)| \Big| = \sqrt{2\,\pi}\,\bar\Phi(t) = \mathrm{o}(\exp(-t^2/2))$$
$$V^2(t) = b^2 - (b + \mathrm{o}(\exp(-t^2/2)))^2 = \mathrm{o}(\exp(-t^2/2))$$

For $\rho(t)$ and $\kappa(t)$, we introduce

$$M(t) := \mathrm{E}[\psi(X-t)^3], \qquad N(t) := \mathrm{E}[\psi(X-t)^4]$$

Then, again suppressing the argument t

$$\rho = V^{-3}[M - 3LS + 2L^3], \qquad \kappa = V^{-4}[N - 4ML + 6SL^2 - 3L^4] - 3$$

and hence

$$\rho_0 = v_0^{-3} M(0), \qquad \kappa_0 = V^{-4} N(0) - 3$$

For ρ_1 we note

$$\rho' = V^{-3}\Big(-3[M - 3LS + 2L^3]\, V'/V + (M' - 3L'S - 3LS' + 3L'L^2) \Big)$$

so that

$$\rho_1 = v_0^{-3}(-3M(0)\tilde v_1 + M'(0) + 3S(0))$$

In our case, for $c = \infty$,

$$M(t) = -3t - t^3, \qquad M'(t) = -3 - 3t^2, \qquad N(t) = t^4 + 6t^2 + 3$$

and for $c = 0$

$$M(t) = (\sqrt{\frac{\pi}{2}}\,)^3 (1 - 2\Phi(t)), \qquad M'(t) = -2\,(\sqrt{\frac{\pi}{2}}\,)^3 \varphi(t), \qquad N(t) = \frac{\pi^2}{4}$$

for $0 < c < \infty$

$$\begin{aligned}
M(t) &= A_c^3\Big[c^3 - \Phi(t+c)(c^3+t^3+3t) - \Phi(t-c)(c^3-t^3-3t) + \\
&\qquad +(t^2+tc+2+c^2)\varphi(t-c) - (t^2-tc+c^2+2)\varphi(t+c)\Big] \\
M'(t) &= A_c^3\Big[3\big(\Phi(t-c)-\Phi(t+c)\big)(t^2+1) - \\
&\qquad -3(t-c)\varphi(t+c) + 3(t+c)\varphi(t-c)\Big] \\
N(t) &= A_c^4\Big[c^4 + \big(\Phi(t+c)-\Phi(t-c)\big)(t^4+6t^2+3-c^4) + \\
&\qquad +(t^3-t^2c+tc^2-c^3+5t-3c)\varphi(t+c) - \\
&\qquad -(t^3+t^2c+tc^2+c^3+5t+3c)\varphi(t-c)\Big]
\end{aligned}$$

This gives the assertion as to ρ_0, ρ_1 and κ_0, and (3.7) and (3.8) also hold.
For $c > 0$, $\Pr(|\eta_c| < b) > 0$ and η_c is continuous. But, on $\{|\eta_c| < b\}$, $\mathcal{L}(\eta_c)$ is a.c. and hence by Lemma 3.2 (C) holds. ////

8.4 Proof of Theorem 3.6

We plug in $(X_i) \sim Q_n$ for some $Q_n \in \tilde{\mathcal{Q}}_n(r)$ into the defining relations for M-estimators of (1.13).

8.4.1 Outline of the proof

We begin with conditioning w.r.t. the number $K = \sum_i U_i = k$ of contaminated observations; next for fixed $t \in \mathbb{R}$, we consider $\tilde{T}_{n,k,t}(t) = \sum_{i:U_i=1} \psi(X_i - t)$ and condition the probability w.r.t. its realization $\tilde{t}_{n,k,t}$. In the sequel we suppress the indices of $\tilde{t}_{n,k,t}$. Denote this event by

$$D_{k,\tilde{t}} := \{K = k, \tilde{T}_{n,k}(\sqrt{t}\,) = \tilde{t}\,\} \tag{8.9}$$

Thus

$$\begin{aligned}
n\,\mathrm{MSE}(S_n, Q_n \,|\, D_{k,\tilde{t}}\,) &= \int_0^\infty \Pr(S_n^2 \geq t \,|\, D_{k,\tilde{t}}\,)\,dt = \\
= \int_0^\infty \Pr(S_n \geq \sqrt{t} \,|\, D_{k,\tilde{t}}\,)\,dt &+ \int_0^\infty \Pr(S_n \leq -\sqrt{t} \,|\, D_{k,\tilde{t}}\,)\,dt
\end{aligned} \tag{8.10}$$

For the sequel, we define

$$\bar{n} := n - k, \qquad s_{n,k} := s_{n,k}(t) = \frac{-\tilde{t} - \bar{n}L(t)}{\sqrt{\bar{n}}\;V(t)} \tag{8.11}$$

To derive the result, we then partition the integrand according to the following tableau where $C' > 0$ is some constant and δ is the exponent from assumption (Vb):

<table>
<tr><th></th><th>$K < k_1 r\sqrt{n}$</th><th>$k_1 r\sqrt{n} \le K < \varepsilon_0 n$</th><th>$K \ge \varepsilon_0 n$</th></tr>
<tr><td>$|t| \le k_2 b^2 \log(n)/n$</td><td>(I)</td><td rowspan="2">(II)</td><td rowspan="3">excluded</td></tr>
<tr><td>$k_2 b^2 \log(n)/n < |t| \le Cn^{1+3/\delta}$</td><td>(III)</td></tr>
<tr><td>$|t| > Cn^{1+3/\delta}$</td><td colspan="2">(IV)</td></tr>
</table>

At this point we also summarize the constants that will be used throughout this section.

constant	k_1	k_2
value	> 1	$> 2 \vee (\frac{3}{2} + \frac{3}{2\delta})$

For all cases except for (I), we will show that they contribute only terms of order $\mathrm{o}(n^{-1})$ to $n\,\mathrm{MSE}(S_n)$ and hence can be neglected. Applying Taylor expansions at large, we derive an expression in which it becomes clear, that independently from t and eventually in n, the maximal MSE is attained for $\tilde{t}_{n,k}$ either kb or identically $-kb$ for all t in (I) — or equivalently all contaminated observations are either smaller than $\check{y}_n - k_2 b^2 \log(n)/n$ or larger than $\hat{y}_n + k_2 b^2 \log(n)/n$. Integrating out first t and then k we obtain the result (3.20) stated in Theorem 3.6.

8.4.2 Conditioning w.r.t. the number of contaminated observations

As announced, for the moment we condition w.r.t. the number $K = \sum_i U_i = k$ of contaminated observations in the sample. Denote the ideally distributed part as $T_{n,k}(t) := \sum_{i:U_i=0} \psi_t(X_i)$. Then we get

$$\Pr\{S_n \le t \,\Big|\, K = k\,\} + R_n^{(0)}(k) = \Pr(T_{n,k}(t) < -\tilde{T}_{n,k}(t)) =$$
$$= \Pr\Big(\frac{T_{n,k}(t) - \bar{n}L(t)}{\sqrt{\bar{n}}V(t)} < -\frac{\tilde{T}_{n,k}(t) - \bar{n}L(t)}{\sqrt{\bar{n}}V(t)}\Big) \tag{8.12}$$

where $R_n^{(0)}(k) \neq 0$ can only happen for mass points of $\mathcal{L}(T_{n,k}(t) + \tilde{T}_{n,k}(t))$.

8.4.3 Conditioning w.r.t. the actual contamination

Next, we condition the probability w.r.t. the actual value of the contamination $\tilde{T}_{n,k} = \tilde{t}$. This gives

$$\Pr\{S_n \le t\,|D_{k,\tilde{t}}\} + \tilde{R}_n^{(0)}(k,\tilde{t}) = \Pr\Big(\frac{T_{n,k}(t) - \bar{n}L(t)}{\sqrt{\bar{n}}\,V(t)} < s_{n,k}(t)\Big) \tag{8.13}$$

where again $\tilde{R}_n^{(0)}(k,\tilde{t}) \neq 0$ can only happen for mass points of $\mathcal{L}(T_{n,k}(t))$.

8.4.4 Negligibility of case (IV)

Without loss, assume that $b = \hat{b}$. By monotonicity and boundedness in assumption (bmi), to given $0 < \eta < -\check{b}$ there is a $t_0 > 0$ such that for $t > t_0$,

$$\check{b} < L(t) = \mathrm{E}[\psi(X^{\mathrm{id}} - t)] \le \check{b} + \eta$$

Let $t_1 > t_0$, $\delta > 0$ and $C' > 0$ so that for $t > t_1$, by (Vb), $|V(t)| \le C' t^{-1-\delta}$. Then we apply the Chebyshev inequality to obtain for $t > t_1^2$

$$\begin{aligned}
&\Pr\{S_n > \sqrt{t}\,\big|\, D_{k,\tilde t}\} \stackrel{(8.2)}{\le} \Pr\Big(T_{n,k}(\sqrt{t}\,) - \bar n L(\sqrt{t}\,) \ge -\tilde t - \bar n L(\sqrt{t}\,)\Big) \le \\
\stackrel{\text{Cheb.}}{\le}\;& \frac{\bar n V^2(\sqrt{t}\,)}{(\tilde t + \bar n L(\sqrt{t}\,))^2} \stackrel{(\text{Vb})}{\le} \frac{\bar n C' t^{-(1+\delta)}}{(\tilde t + \bar n L(\sqrt{t}\,))^2} \le \frac{n C' t^{-(1+\delta)}}{(\tilde t + \bar n \check b + \eta)^2} \le \\
\stackrel{\tilde t \le k\hat b}{\le}\;& \frac{n C' t^{-(1+\delta)}}{[k\hat b + \bar n \check b + \eta]^2} = \frac{n C' t^{-(1+\delta)}}{[k(\hat b - \check b) + n\check b + \eta]^2} \stackrel{k \le \varepsilon_0 n}{\le} \frac{n C' t^{-(1+\delta)}}{(\check b - \eta)^2}
\end{aligned} \tag{8.14}$$

and correspondingly (with $b = -\check b$) for $\Pr\{S_n \le -\sqrt{t}\,\big|\, D_{k,\tilde t}\}$; but

$$\frac{C' n^2}{(b-\eta)^2} \int_{Cn^{1+3/\delta}}^{\infty} t^{-(1+\delta)}\, dt = \frac{C' C^{-\delta} n^{-1-\delta}}{\delta(\check b - \eta)^2} = \mathrm{o}(n^{-1}) \tag{8.15}$$

8.4.5 Negligibility of case (II)

For the proof of Theorem 3.6, a weaker version of the following lemma, Ruckdeschel (2005a, Lemma 5.3), would suffice to settle case (II), but for the proof of Theorem 6.5, we have to allow for k_1 varying in n.

Lemma 8.1 *Let $k_1(n) = 1 + d_n$ and assume that for some $\delta \in (0, 1/4)$,*

$$d_n n^{1/4-\delta} \to \infty, \qquad d_n n^{-1/4+\delta} \to 0 \qquad \text{for } n \to \infty \tag{8.16}$$

Let

$$\mathcal{K}_n := k_1(n) \log k_1(n) + 1 - k_1(n) \tag{8.17}$$

Then if $\liminf_n d_n > 0$ there is some $c > 0$ such that

$$\Pr(\mathrm{Bin}(n, r/\sqrt{n}\,) > k_1(n) r \sqrt{n}\,) = \mathrm{o}(e^{-cr\sqrt{n}}) \tag{8.18}$$

and, if $d_n = \mathrm{o}(n^0)$, for any $0 < \delta_0 \le 2\delta$, it holds that

$$\Pr(\mathrm{Bin}(n, r/\sqrt{n}\,) > k_1(n) r \sqrt{n}\,) = \mathrm{o}(e^{-rn^{\delta_0}}) \tag{8.19}$$

Remark 8.2 Even if d_n is increasing at a faster rate than $n^{1/4}$, assertion (8.18) remains true, as long as $\liminf_n d_n > 0$ —but this is not needed here.

Proof : We first note that $\mathcal{K}_n > 0$, as $\log(x) > 0$ for $x > 1$ and

$$\mathcal{K}_n = \int_1^{k_1(n)} \log(x)\, dx \tag{8.20}$$

Applying Hoeffding's Lemma 9.2 to the case of n independent $\mathrm{Bin}(1, p)$ variables, we obtain for $B_n \sim \mathrm{Bin}(n, p_n)$, $p_n = r/\sqrt{n}$ and $\varepsilon = (k_1(n) - 1)r/\sqrt{n}$ (which is smaller than $1 - p_n$ eventually)

$$\begin{aligned}
\Pr(B_n > k_1(n) r\sqrt{n}\,) \quad \le \quad & \exp\Big(- k_1(n) r \sqrt{n} \log(k_1(n)) + (n - k_1(n) r \sqrt{n}\,) \times \\
& \times \big(\log(1 - \frac{r}{\sqrt{n}}) - \log(1 - k_1(n)\frac{r}{\sqrt{n}})\big)\Big)
\end{aligned}$$

But for $x \in (0,1)$, $-\frac{x}{1-x} \le \log(1-x) \le -x$. Thus

$$\log(1-\frac{r}{\sqrt{n}}) - \log(1-k_1(n)\frac{r}{\sqrt{n}}) \le \frac{k_1(n)r}{\sqrt{n}\,(1-k_1(n)r/\sqrt{n}\,))} - \frac{r}{\sqrt{n}}$$

$$\begin{aligned}\Pr(B_n > k_1(n)r\sqrt{n}\,) &\le \exp\Big(-r\sqrt{n}\,\big(k_1(n)\,\log(k_1(n)) - k_1(n)+1\big) + rk_1(n)^2\Big) = \\ &= \exp\big(-\mathcal{K}_n\,r\sqrt{n} + rR_n\big)\end{aligned}$$

for $R_n = \mathrm{O}(1) + \mathrm{O}(d_n^2)$, where due to the second assumption in (8.16), $d_n^2 = \mathrm{o}(\sqrt{n}\,)$. If $\liminf_n d_n > 0$, by (8.20) $\liminf_n \mathcal{K}_n > 0$, and for any $0 < c < \liminf_n \mathcal{K}_n$, (8.18) follows. If $d_n = \mathrm{o}(n^0)$, we note that

$$\mathcal{K}_n = (1+d_n)\log(1+d_n) - d_n = d_n^2/2 + \mathrm{o}(d_n^2) \tag{8.21}$$

which for any $\delta' > 0$ entails

$$\Pr(\mathrm{Bin}(n, r/\sqrt{n}\,) > k_1(n)r\sqrt{n}\,) = \mathrm{o}\Big(\exp\big(-\frac{rd_n^2\sqrt{n}}{2+\delta'}\big)\Big)$$

Now for $d_n = \mathrm{o}(n^0)$, by the first assumption in (8.16), for $0 < \delta_0 < 2\delta$ eventually in n, (8.19) holds as

$$n^{\delta_0} - \frac{d_n^2\sqrt{n}}{2+\delta'} < n^{2\delta}(1 - \frac{n^{1/2-2\delta}d_n^2}{2+\delta'}) \to -\infty$$

////

As in (II), $|t| < Cn^{1+3/\delta}$, the integrand of $n\,\mathrm{MSE}(S_n, Q_n \mid D_{k,\tilde t}\,)$ is bounded by some polynomial in n, and hence by Lemma 8.1 the contribution of (II) is indeed $\mathrm{o}(n^{-1})$.

Another consequence of the exponential decay of (8.18)/(8.19) is that we may neglect values of $K > k_1(n)r\sqrt{n}$ when integrating along K.

Corollary 8.3 *Let* $K \sim \mathrm{Bin}(n, r/\sqrt{n}\,)$*. Then, in the setup of Lemma 8.1, for any* $j \in \mathbb{N}$,

$$\mathrm{E}[K^j\,\mathrm{I}_{\{X \ge k_1(n)r\sqrt{n}\}}] = \mathrm{o}(e^{-rn^d}) \tag{8.22}$$

for any $0 < d < \sqrt{n}$ *if* $\liminf_n d_n > 0$ *and any* $0 < d \le \delta_0$ *if* $\lim_n d_n = 0$.

PROOF : $\mathrm{E}[K^j\,\mathrm{I}_{\{K \ge k_1(n)r\sqrt{n}\}}] \le n^j \Pr(X > k_1(n)r\sqrt{n}) \overset{(8.18)/(8.19)}{=} \mathrm{o}(e^{-rn^d})$ ////

8.4.6 Negligibility of case (III)

We apply Hoeffding's bound Lemma 9.1:

$$\Pr\{S_n > \sqrt{t}\,\Big|\, D_{k,\tilde t}\} \le \Pr(T_{n,k}(\sqrt{t}\,) \ge -\tilde t\,\big|\, D_{k,\tilde t}\,) \le \exp(-2n\Delta^2/b^2) \tag{8.23}$$

for $\Delta := -L(\sqrt{t}\,) - \frac{\tilde t}{n}$. As ψ is isotone, L is antitone, hence in case (III),

$$L(\sqrt{t}\,) \le L(b\sqrt{k_2\log(n)/n}\,) = -b\sqrt{k_2\log(n)/n} + \mathrm{o}(\sqrt{\log(n)/n}\,) \qquad (8.24)$$

Thus

$$\Delta \ge -L(\sqrt{t}\,) - \frac{kb}{n} \stackrel{(8.24)}{>} \frac{b}{\sqrt{n}}[\sqrt{k_2\log(n)} + \mathrm{o}(\sqrt{\log(n)}\,)] \qquad (8.25)$$

and

$$\exp(-2\frac{n\,\Delta^2}{b^2}) < n^{-2k_2}(1+\mathrm{o}(n^0)) \qquad (8.26)$$

This latter is $\mathrm{o}(n^{-3-3/\delta})$ and thus integrating $n\,\mathrm{MSE}$ out along (III) we get something of order $\mathrm{o}(n^{-1})$.

8.4.7 Asymptotic normality

On (I), by Lemma 1.1

$$\Pr\big\{S_n \ge \sqrt{t}\,\big|\, D_{k,\tilde t}\,\big\} = \Pr\left(\frac{T_{n,k}(\sqrt{t}\,) - \bar n L(\sqrt{t}\,)}{\sqrt{\bar n}\,V(\sqrt{t}\,)} > s_{n,k}(t)\right) + \mathrm{O}(e^{-\gamma n}) \qquad (8.27)$$

for some $\gamma > 0$, uniformly in t and k. For $i = 1,\ldots,\bar n$, let $j_i \in \{1,\ldots,n\}$ be the indices such that $U_{j_i} = 0$. We may apply Theorem 9.3(b) to (8.10)/(8.13), identifying

$$\xi_{i,t} := \frac{1}{V(t)}[\psi_t(X_{j_i}) - L(t)], \qquad i = 1,\ldots,\bar n \qquad (8.28)$$

and setting $\Theta := \Theta_n = \{|t| \le k_2 b^2\log(n)/n\}$. This application is possible, as $|\psi| < b$, so $\sup_{t\in\Theta_n}\mathrm{E}\,|\tilde\xi_{i,t}|^5 < \infty$. By condtion (C) of our assumptions, Cramér condition (9.11) of the theorem holds if n is large enough.
We note that if in Theorem 3.6, we limit ourselves to term A_1 and hence only assume (C'), we may apply Theorem 9.3(a).
With $G_{n,t}(s)$ from (9.7) we define

$$\tilde G_{n,t}(u) := G_{n,t}(s_{n,k}(u)), \qquad \tilde G_n(t) := \tilde G_{n,t}(t) \qquad (8.29)$$

With these definitions we have for $|t| \le k_2 b^2\log(n)/n$ and $K < k_1 r\sqrt{n}$ uniformly in t and k:

$$\begin{aligned} &\mathrm{O}(\exp(-\gamma n)) + \Pr\big\{S_n \ge \sqrt{t}\,\big|\, D_{k,\tilde t}\,\big\} = \\ = \quad &\Pr\big(\sum_{i=1}^{\bar n}\xi_{i,\sqrt{t}} > s_{n,k}(\sqrt{t}\,)\big) = 1 - \tilde G_n(\sqrt{t}\,) + \mathrm{O}(n^{-3/2}) \end{aligned} \qquad (8.30)$$

Hence, using negligibility of (II), (III) and (IV), and setting

$$n^\natural = \sqrt{\bar n/n}, \qquad l_n = n^\natural\sqrt{k_2\log(n)}, \qquad l_n^{(0)} = k_2 b^2\log(n)/n \qquad (8.31)$$

we obtain

$$n\,\mathrm{MSE}(S_n,Q_n\,\big|\,D_{k,\tilde t}\,) = (n^\natural)^{-2}\,\bar n\,\int_0^{l_n^{(0)}} 1-\tilde G_n(\sqrt{t}\,)+\tilde G_n(-\sqrt{t}\,)\,dt+\mathrm{o}(n^{-1}) =$$

$$= \quad 2(n^\natural)^{-2}\int_0^{bl_n} u\Big(1-\tilde G_n(\frac{u}{\sqrt{\bar n}}\,)+\tilde G_n(-\frac{u}{\sqrt{\bar n}}\,)\Big)\,du+\mathrm{o}(n^{-1}) \qquad (8.32)$$

As $\tilde G_n$ is arbitrarily smooth, integration by parts is available and gives

$$n\,\mathrm{MSE}(S_n,Q_n\,\big|\,D_{k,\tilde t}\,) = R_n+(n^\natural)^{-2}\int_{-bl_n}^{bl_n}\frac{u^2}{\sqrt{\bar n}}\;G_n'(\frac{u}{\sqrt{\bar n}}\,)\,du+\mathrm{o}(n^{-1}) \qquad (8.33)$$

with

$$R_n := k_2\,\log(n)\,b^2\,\big[1-\tilde G_n(b\sqrt{\tfrac{k_2\log(n)}{n}}\,)-\tilde G_n(-b\sqrt{\tfrac{k_2\log(n)}{n}}\,)\,\big] \qquad (8.34)$$

A closer look at $s_{n,k}(\pm b\sqrt{\tfrac{k_2\log(n)}{n}}\,)$ reveals

$$s_{n,k}(\pm b\sqrt{\tfrac{k_2\log(n)}{n}}\,) \quad \stackrel{(3.6)}{=} \quad \frac{\mathrm{O}(\sqrt{n}\,)\pm b\sqrt{\frac{k_2\bar n^2\log(n)}{n}}+\mathrm{O}(\frac{\bar n\,\log(n)}{n}\,)}{\sqrt{\bar n}\,(v_0+\mathrm{o}(n^0))} =$$

$$= \quad \frac{\pm b\sqrt{k_2\log(n)}}{v_0}(1+\mathrm{o}(n^0)) \qquad (8.35)$$

We also note that, again by (bmi) $v_0^2=\mathrm{E}[\psi^2]\le b^2$, hence $b/v_0>1$. In particular, eventually in n,

$$|\tilde s_{n,k}(\pm b\sqrt{k_2\log(n)}\,)|>\sqrt{2\log(n)} \qquad (8.36)$$

But, as $|\psi|\le b$ by (bmi), $|\kappa|\le b^4$ and $|\rho|\le b^3$, and thus by Mill's ratio, there is some $0<K<\infty$, independent of t, n, such that for any $s>0$

$$\max\big(1-G_{n,t}(s),\,G_{n,t}(-s)\big)\le K|s|^5\,\exp(-s^2/2) \qquad (8.37)$$

Thus for n sufficiently large

$$1-\tilde G_n(b\sqrt{\tfrac{k_2\log(n)}{n}}\,)=\exp(-\frac{k_2b^2\log(n)}{2v_0^2}+\mathrm{o}(n^0)))=\mathrm{O}(\frac{\log(n)^{5/2}}{n^{1+\delta}}) \qquad (8.38)$$

for some $\delta>0$. The same goes for $\tilde G_n(-2b\sqrt{\tfrac{\log(n)}{n}}\,)$, and therefore,

$$R_n=\mathrm{O}(\log(n)^{7/2}/n^{1+\delta})=\mathrm{o}(n^{-1}) \qquad (8.39)$$

and

$$n\,\mathrm{MSE}(S_n,Q_n\,\big|\,D_{k,\tilde t}\,)=(n^\natural)^{-2}\int_{-bl_n}^{bl_n}\frac{u^2}{\sqrt{\bar n}}\;G_n'(\frac{u}{\sqrt{\bar n}}\,)\,du+\mathrm{o}(n^{-1}) \qquad (8.40)$$

To make more transparent, which terms are bounded to which degree, we introduce the following notation, which will also help `MAPLE` to ignore irrelevant terms

$$t^\natural := \frac{\tilde t}{\sqrt{\bar n}}, \qquad \tilde s_{n,k}(x) = s_{n,k}(\frac{x}{\sqrt{\bar n}}), \tag{8.41}$$

Then on (I), $u = \mathrm{O}(\sqrt{\log(n)})$, $t^\natural = \mathrm{O}(n^0)$. In particular this will not affect the remainder terms of the Taylor expansions of assumption (D).
In the sequel, we drop the indices of $s_{n,k}$ and $\tilde s_{n,k}$, where they are clear from the context. Next, we spell out $\tilde G_n'(u)$ in (8.40) more explicitly. Denote

$$\mathcal{G}_n(s,t) := G_{n,t}(s), \quad G^{(1)}_{n,t}(s) := [\tfrac{\partial}{\partial s}\mathcal{G}_n](s,t), \quad G^{(2)}_{n,t}(s) := [\tfrac{\partial}{\partial t}\mathcal{G}_n](s,t) \tag{8.42}$$

Then, as $\tilde s'_{n,k}(x) = s'_{n,k}(\frac{x}{\sqrt{\bar n}})/\sqrt{\bar n}$,

$$\begin{aligned} \tilde G_n'(\tfrac{u}{\sqrt{\bar n}}) &= [G^{(1)}_{n,x}(s(x))s'(x) + G^{(2)}_{n,x}(s(x))]\Big|_{x=\frac{u}{\sqrt{\bar n}}} = \\ &= G^{(1)}_{n,u/\sqrt{\bar n}}(\tilde s(u))\,\tilde s'(u)\sqrt{\bar n} + G^{(2)}_{n,u/\sqrt{\bar n}}(\tilde s(u)) =: \tilde g_n(u)\sqrt{\bar n} \end{aligned} \tag{8.43}$$

and therefore

$$n\,\mathrm{MSE}(S_n, Q_n \,\big|\, D_{k,\tilde t}) = (n^\natural)^{-2}\int_{-bl_n}^{bl_n} u^2 \tilde g_n(u)\,du + \mathrm{o}(n^{-1}) \tag{8.44}$$

8.4.8 Expanding $\tilde g_n(u)$

Considering $\tilde g_n(u)$ more closely, we expand the terms according to assumption (D) —with the help of our `MAPLE` procedures `asS`, `asS1`, `asg`

$$\begin{aligned} \tilde s(u) = \frac{-t^\natural - \sqrt{\bar n}L(\frac{u}{\sqrt{\bar n}})}{V(\frac{u}{\sqrt{\bar n}})} &= \tfrac{1}{v_0}\Big[(u-t^\natural) - \tfrac{u}{\sqrt{\bar n}}\big(\tfrac{l_2 u}{2} + \tilde v_1(u-t^\natural)\big) + \\ &+ \tfrac{1}{\bar n}\Big((l_2\tfrac{\tilde v_1}{2} - \tfrac{l_3}{6})u^3 + (u-t^\natural)u^2(\tilde v_1^2 - \tilde v_2/2)\Big)\Big] + \mathrm{O}(n^{-(1+\delta)}) \end{aligned} \tag{8.45}$$

$$\begin{aligned} \tilde s'(u) = -\frac{L'(\frac{u}{\sqrt{\bar n}})}{V(\frac{u}{\sqrt{\bar n}})} + \frac{(t^\natural + L(\frac{u}{\sqrt{\bar n}}))V'(\frac{u}{\sqrt{\bar n}})}{V^2(\frac{u}{\sqrt{\bar n}})} &= \tfrac{1}{v_0}\Big[1 - l_2\tfrac{u}{\sqrt{\bar n}} - 2\tilde v_1\tfrac{u}{\sqrt{\bar n}} + \tfrac{t^\natural}{\sqrt{\bar n}}\tilde v_1 + \\ &+ \tfrac{1}{\bar n}\Big((3\tilde v_1^2 - \tfrac{l_3}{2} - \tfrac{3}{2}\tilde v_2 + \tfrac{3}{2}\tilde v_1 l_2)u^2 + ut^\natural(\tilde v_2 - 2\tilde v_1^2)\Big)\Big] + \mathrm{O}(n^{-(1+\delta)}) \end{aligned} \tag{8.46}$$

as well as

$$\begin{aligned} G^{(1)}_{n,u/\sqrt{\bar n}}(\tilde s) &= \varphi(\tilde s)\Big[1 + \tfrac{1}{6\sqrt{\bar n}}(\rho_0 + \rho_1\tfrac{u}{\sqrt{\bar n}})(\tilde s^3 - 3\tilde s) + \tfrac{1}{24n}\kappa_0(\tilde s^4 - 6\tilde s^2 + 3) + \\ &+ \tfrac{1}{72n}\rho_0^2(\tilde s^6 - 15\tilde s^4 + 45\tilde s^2 - 15)\Big] + \mathrm{O}(n^{-(1+\delta)}) \end{aligned} \tag{8.47}$$

and respectively,

$$G^{(2)}_{n,u/\sqrt{\bar n}}(\tilde s) = \varphi(\tilde s)\tfrac{\rho_1}{6\sqrt{\bar n}}(1 - \tilde s^2) + \mathrm{O}(n^{-(1/2+\delta)}) \tag{8.48}$$

This gives

$$\tilde g_n(u) = v_0\varphi(\tilde s)[1+\tfrac{1}{\sqrt{\tilde n}}P_1(u,t^\natural)+\tfrac{1}{\tilde n}P_2(u,t^\natural)]+\mathrm{O}(n^{-(1+\delta)}) \tag{8.49}$$

for

$$P_1(u,t^\natural) = -l_2u-2\tilde v_1u+t^\natural\tilde v_1+\tfrac{\rho_0}{6v_0^3}(u-t^\natural)^3-\tfrac{\rho_0}{2v_0}(u-t^\natural) \tag{8.50}$$

and $P_2(u,t^\natural)$ a corresponding polynomial in u, $t^\natural$, $\tilde v_1$, $\tilde v_2$, l_2, l_3, ρ_0, ρ_1, and κ_0, the exact expression of which may be taken from our `MAPLE` procedure `asg`.
To be able to calculate the integrals, we expand $\varphi(\tilde s)$ in a Taylor expansion about

$$s_1=(u-t^\natural)/v_0 \tag{8.51}$$

as

$$\varphi(\tilde s)=\varphi(s_1)[1-s_1(\tilde s-s_1)+(s_1^2-1)(\tilde s-s_1)^2/2]+\mathrm{O}(n^{-(1+\delta)}) \tag{8.52}$$

and hence

$$\tilde g_n(u)=v_0\varphi(s_1)g_n(s_1)+\mathrm{O}(n^{-(1+\delta)}) \tag{8.53}$$

with

$$g_n(s_1):=1+\tfrac{1}{\sqrt{\tilde n}}\tilde P_1(s_1,t^\natural)+\tfrac{1}{\tilde n}\tilde P_2(s_1,t^\natural) \tag{8.54}$$

for

$$\begin{aligned}\tilde P_1(s_1,t^\natural) \quad = \quad & \rho_0\frac{s_1^3-3s_1}{6}+(\tfrac{l_2}{2}+\tilde v_1)s_1^3-(l_2+2\tilde v_1)s_1v_0+\\ & +(l_2+\tilde v_1)[s_1^2-1]t^\natural+\frac{(t^\natural)^2l_2s_1}{2v_0}\end{aligned} \tag{8.55}$$

and $\tilde P_2(s_1,t^\natural)$ a corresponding polynomial again to be looked up from our `MAPLE` procedure `asgns`. This gives

$$n\,\mathrm{MSE}(S_n,Q_n\,\big|\,D_{k,\tilde t}\,)=(n^\natural)^{-2}\int_{-bl_n/v_0}^{bl_n/v_0}h_n(s)\varphi(s)\,\lambda(ds)+\mathrm{o}(n^{-1}) \tag{8.56}$$

for

$$h_n(s)=u_1(s)^2g_n(s),\qquad u_1(s)=sv_0+t^\natural \tag{8.57}$$

8.4.9 Selection of the least favorable contamination

Function $h_n(s)$ from (8.57) is a polynomial in s, hence on (I), where $|s|=\mathrm{O}(\log(n))$, we may ignore terms of (pointwise-in-s) order $\mathrm{O}(n^{-(1+\delta)})$. This gives a complicated expression of form

$$h_n(s)=(sv_0+t^\natural)^2+\frac{1}{\sqrt{n}}Q_1+\frac{1}{n}Q_2 \tag{8.58}$$

where v_0Q_1 is a polynomial in s, $t^\natural$, v_0, l_2, $\tilde v_1$, and ρ_0 with $\deg(Q_1,s)=5$ and $\deg(Q_1,t)=4$, and $v_0^2Q_2$ is a polynomial in s, $t^\natural$, v_0, l_2, $\tilde v_1$, ρ_0, l_3, $\tilde v_2$, $\tilde\rho_1$, and

κ_0 with $\deg(Q_2, s) = 8$ and $\deg(Q_1, t) = 6$; the exact expressions are available on the web-page and may be generated by our `MAPLE`-procedure `ashn`. Denoting the second partial derivative w.r.t. $t^\natural$ by an index t, t we consider

$$h_{n,t,t}(s) = 2 + \frac{1}{\sqrt{n}} Q_{1,t,t} + \frac{1}{n} Q_{2,t,t} \tag{8.59}$$

where $\deg(Q_{1,t,t}, s) = 3$ and $\deg(Q_{2,t,t}, s) = 6$, and under (7.1), i.e., if $l_2 = \tilde v_1 = \rho_0 = 0$, $Q_{1,t,t} = 0$ and $\deg(Q_{2,t,t}, s) = 4$. That is, on (I), uniformly in s, $h_{n,t,t}(s) = 2 + \mathrm{O}(\log(n)^3/\sqrt{n})$, and under (7.1), the remainder is even $\mathrm{O}(\log(n)^4/n)$. Hence eventually in n, uniformly in s, h_n is strictly convex in $t^\natural$, hence takes its maximum on the boundary, that is for $|t^\natural|$ maximal.

Going back to the definition of $t^\natural$, we note that for fixed n and k,

$$t^\natural = \tilde t/\sqrt{\bar n} = \sum_{i: U_i = 1} \psi(X_i - t)/\sqrt{\bar n} \tag{8.60}$$

Obviously, $\tilde t$ is bounded in absolute value by kb. This value may be attained if (up to $\mathrm{O}(n^{-1})$) all terms $\psi(X_i - t)$ are either b or $-b$ for all t in (I). This amounts to concentrating essentially all the contamination either right of $\hat y_n + b\sqrt{k_2 \log(n)/n}$ or left of $\check y_n - b\sqrt{k_2 \log(n)/n}$; the decision which of the two alternatives is least favorable is deferred to subsubsection 8.4.13.

As we may allow for deviations from this "outlyingness" as long as we do no affect the expansion of the MSE up to $\mathrm{O}(n^{-1})$, we may weaken the concentration property to (3.23) resp. (3.24): On (I), $|t^\natural|$ is bounded, so smallness of the probabilities in (3.23) resp. (3.24) entails that also the expectations of $(t^\natural)^j$, $j = 1, \ldots, 6$ arising in $h_n(s)$ are $\mathrm{o}(n^{-1})$.

Denote a distribution in $\tilde{\mathcal{Q}}_n$ which is contaminated according to (3.23) resp. (3.24) by Q_n^0. By the previous considerations, under Q_n^0, we may consider $|\tilde t|$ as being exactly kb, and we will consider the cases $\tilde t = \pm kb$ simultaneously. For the substitution $t^\natural = \pm kb/\sqrt{\bar n}$, the following abbreviations are convenient

$$\tilde k := k/\sqrt{n}, \qquad k^\natural := k/\sqrt{\bar n} = \tilde k / n^\natural \tag{8.61}$$

Taking up the dependency on $t^\natural$ in $h_n(s)$ as $h_n(s) = h_n(s, t^\natural)$, in the `MAPLE` procedure `ash`, we introduce

$$\tilde h_n(s) = \tilde h_n(s, k^\natural) = h_n(s, k^\natural b) \tag{8.62}$$

8.4.10 Integration w.r.t. s

In this step we integrate out s in $\tilde h_n(s)$. As $bl_n/v_0 > \sqrt{2\log(n)}$, by Lemma 9.5, we may drop the integration limits and get

$$n \,\mathrm{MSE}(S_n, Q_n^0 \,|K = k\,) = (n^\natural)^{-2} \int_{-\infty}^{\infty} \tilde h_n(s) \varphi(s)\, \lambda(ds) + \mathrm{o}(n^{-1}) \tag{8.63}$$

So for integration, we use that for $X \sim \mathcal{N}(0,1)$, $\mathrm{E}[X^j] = 0$, for $j = 1, 3, 5, 7$, and

$$\mathrm{E}[X^2] = 1, \quad \mathrm{E}[X^4] = 3, \quad \mathrm{E}[X^6] = 15, \quad \mathrm{E}[X^8] = 115 \tag{8.64}$$

and get (by our `MAPLE` procedures `intesout` and `asMSEK`)

$$
\begin{aligned}
&n\,\mathrm{MSE}(S_n,Q_n^0\,\big|K=k\,)=\\
=\;\;&\mathrm{o}(n^{-1})+(n^\natural)^{-2}\Big[(k^\natural)^2b^2+v_0^2+\tfrac{1}{\sqrt{\bar n}}[\pm(3l_2+4\tilde v_1)v_0^2k^\natural b\pm l_2(k^\natural)^3b^3]+\\
&\quad+\tfrac{1}{\bar n}\big[(\tfrac54 l_2^2+\tfrac13 l_3)(k^\natural)^4b^4+(3\tilde v_2+2l_3+3\tilde v_1^2+\tfrac{15}{2}l_2+12\tilde v_1 l_2)v_0^2(k^\natural)^2b^2+\\
&\quad+(\rho_0(2\tilde v_1+l_2)+\tfrac23\rho_1)v_0^3)+(12\tilde v_1 l_2+l_3+3\tilde v_2+\tfrac{15}{4}l_2^2+9\tilde v_1^2)v_0^4\big]\Big] \qquad (8.65)
\end{aligned}
$$

As mentioned in Remark 3.7(c), the terms of κ_0 cancel out for A_2 as do the terms of ρ_0 for A_1.

8.4.11 Collection of terms

As we want to calculate the expectation with respect to K, we have to expand terms in a way that k is only appearing in integer powers and in the nominator. For this purpose we employ our `MAPLE` procedures `asNn`, `asKn`, and get

$$
\begin{aligned}
(n^\natural)^{-2} &= 1+\tfrac{\tilde k}{\sqrt n}+\tfrac{\tilde k^2}{n}+\mathrm{o}(n^{-1}) && (8.66)\\
(n^\natural)^{-3} &= 1+\tfrac{3\tilde k}{2\sqrt n}+\mathrm{o}(n^{-1/2}),\qquad (n^\natural)^{-4}=1+\mathrm{o}(n^0) && (8.67)\\
k^\natural &= \tilde k+\tfrac{\tilde k^2}{2\sqrt n}+\mathrm{o}(n^{-1/2}),\qquad (k^\natural)^2=\tilde k^2+\tfrac{\tilde k^3}{\sqrt n}+\tfrac{\tilde k^4}{n}+\mathrm{o}(n^{-1}) && (8.68)\\
(k^\natural)^3 &= \tilde k^3+\tfrac{3\tilde k^4}{\sqrt n}+\mathrm{o}(n^{-1/2}),\qquad (k^\natural)^4=\tilde k^4+\mathrm{o}(n^0) && (8.69)
\end{aligned}
$$

Substituting $k^\natural$ and $n^\natural$ by means of these expressions, we obtain (`MAPLE` procedure `asMSEk`)

$$
\begin{aligned}
&n\,\mathrm{MSE}(S_n,Q_n^0\,\big|\,K=k\,)=\\
=\;&\mathrm{o}(n^{-1})+\tilde k^2b^2+{v_0}^2+\frac{[\pm\,(4\,\tilde v_1\,+3\,l_2\,)\,b+1]\tilde k{v_0}^2+(2\pm l_2b)\,\tilde k^3b^2}{\sqrt n}+\\
&+\frac{\big(3\,b^2\pm 3\,l_2\,b^3+\big(\tfrac54\,{l_2}^2+\tfrac13\,l_3\big)\,b^4\big)\,\tilde k^4}{n}+\\
&+\frac{\big(\big(3\,\tilde v_1^2+3\,\tilde v_2+12\,l_2\tilde v_1+\tfrac{15}{2}\,{l_2}^2+2\,l_3\big)\;b^2+1\pm(6\,l_2\,+8\,\tilde v_1\,)\,b\big)\,\tilde k^2v_0^2}{n}+\\
&+\frac{\big(3\tilde v_2+9\tilde v_1^2+\tfrac{15}{4}\,{l_2}^2+l_3+12\,l_2\,\tilde v_1\big)\,{v_0}^4+\big((l_2+2\tilde v_1)\,\rho_0+\tfrac23\,\rho_1\big)\,{v_0}^3}{n} \qquad (8.70)
\end{aligned}
$$

8.4.12 Integration w.r.t. $\tilde k$

As by Corollary 8.3 the event $\{K>(1+\delta)r\sqrt n\}$ only attributes $\mathrm{o}(n^{-1})$ to the expectation of $\mathrm{E}[K^j]$, $j=0,\ldots,4$, we can now simply use Lemma 8.1 to determine the MSE. This gives the result by our `MAPLE` procedures `intekout`, `asMSE`:

$$
n\ \mathrm{E}_{Q_n^0}[S_n^2\,]=r^2b^2+{v_0}^2+\tfrac{r}{\sqrt n}\,A_1+\tfrac1n\,A_2+\mathrm{o}(n^{-1}) \qquad (8.71)
$$

with

$$A_1 = v_0{}^2 \Big(\pm (4\,\tilde v_1 + 3\,l_2\,)b + 1\Big) + b^2 + [2\,b^2 \pm l_2\,b^3\,]\,r^2 \tag{8.72}$$

$$\begin{aligned} A_2 = {}& v_0{}^3 \Big((l_2 + 2\,\tilde v_1\,)\rho_0 + \tfrac{2}{3}\,\rho_1\Big) + v_0{}^4\,(3\,\tilde v_2 + \tfrac{15}{4}\,l_2{}^2 + l_3 + 9\,\tilde v_1^2 + 12\,\tilde v_1\,l_2\,) + \\ &+[\,v_0{}^2 \Big((3\,\tilde v_2 + 3\,\tilde v_1^2 + \tfrac{15}{2}\,l_2{}^2 + 2\,l_3 + 12\,\tilde v_1\,l_2\,)b^2 + 1 \pm (8\,\tilde v_1 + 6\,l_2\,)\,b\Big) + \\ &\pm 3\,l_2\,b^3 + 5\,b^2\,]\,r^2 + \Big((\tfrac{5}{4}\,l_2{}^2 + \tfrac{1}{3}\,l_3\,)b^4 \pm 3\,l_2\,b^3 + 3\,b^2\Big)\,r^4 \end{aligned} \tag{8.73}$$

8.4.13 Decision upon the alternative (3.23) or (3.24)

Denote Q_n^- a contaminated member in $\tilde{\mathcal Q}_n(r)$ according to (3.23) and correspondingly Q_n^+ according to (3.24). With respect to terms of (8.71)–(8.73), obviously, if $\sup\psi < -\inf\psi$, the maximal MSE is achieved by Q_n^-, respectively by Q_n^+ if $\sup\psi > -\inf\psi$. In case $\sup\psi = -\inf\psi$, the terms in A_1 are decisive:

$$\begin{aligned} n(\mathrm{E}_{Q_n^+}[S_n^2\,] - \mathrm{E}_{Q_n^-}[S_n^2\,]) = {}& \tfrac{rb}{\sqrt n}\Big\{ l_2\big[(r^2b^2 + 3v_0^2)(1 + 2\tfrac{r}{\sqrt n}) + \tfrac{3b^2r(r^2+1)}{\sqrt n}\big] + \\ &+4v_0^2(1 + \tfrac{r}{\sqrt n})v_1\Big\} + \mathrm{o}(n^{-1}) \end{aligned} \tag{8.74}$$

Hence, Q_n^- $[Q_n^+]$ is least favorable up to $\mathrm{o}(n^{-1})$ if

$$\tilde v_1 \;>\; [<] - \tfrac{l_2}{4}\Big(\tfrac{b^2}{v_0^2}(r^2+3)(1 + \tfrac{r}{\sqrt n} - \tfrac{2r^2}{n}) + 3(1 - \tfrac{b^2}{v_0^2})\Big) \tag{8.75}$$

If there is “$=$” in (8.75), no decision can be taken up to order $\mathrm{o}(n^{-1})$.

8.4.14 Special cases

Obviously, under symmetry, or more exactly under (7.1), we have

$$\begin{aligned} n\;\mathrm{E}_{Q_n^0}[S_n^2\,] = {}& \big(r^2b^2 + v_0{}^2\big)\Big(1 + \tfrac{r}{\sqrt n} + \tfrac{r^2}{n}\Big) + \tfrac{r}{\sqrt n}\;\big(b^2(1+r^2)\big) + \\ &+\tfrac{r^2}{n}\;\big(b^2(5+2r^2)\big) + \frac{\tfrac{2}{3}v_0{}^3\,\rho_1 + v_0{}^4\,(3\,\tilde v_2 + l_3\,)}{n} + \\ &+\frac{\big(v_0{}^2\,(3\,\tilde v_2 + 2\,l_3\,)\,b^2\big)\;r^2 + \tfrac{1}{3}\,l_3 b^4\,r^4}{n} + \mathrm{o}(n^{-1}), \end{aligned} \tag{8.76}$$

and under $r = 0$, we get

$$\begin{aligned} n\;\mathrm{E}_{F^n}[S_n^2\,] = {}& v_0{}^2 + \frac{v_0{}^3\;\big((l_2 + 2\,\tilde v_1\,)\rho_0 + \tfrac{2}{3}\,\rho_1\big)}{n} + \\ &+\frac{v_0{}^4\,(3\,\tilde v_2 + l_3 + \tfrac{15}{4}\,l_2{}^2 + 12\,\tilde v_1\,l_2 + 9\,\tilde v_1^2\,)}{n} + \mathrm{o}(n^{-1}) \end{aligned} \tag{8.77}$$

respectively, again under (7.1),

$$n\;\mathrm{E}_{F^n}[S_n^2\,] = v_0{}^2 + \frac{\tfrac{2}{3}\,v_0{}^3\rho_1 + v_0{}^4\,(3\,\tilde v_2 + l_3\,)}{n} + \mathrm{o}(n^{-1}) \tag{8.78}$$

////

8.5 Proofs to Propositions 6.1 and 6.2

For $\varepsilon_1 \in (0,1)$, let

$$N_+(t) = N_+(t;n,\varepsilon_1,\hat{b}) \quad := \quad \#\Big\{\psi(x_i - t) \geq \hat{b}(1-\varepsilon_1),\ U_i = 0\Big\} \tag{8.79}$$

$$N_-(t) = N_-(t;n,\varepsilon_1,\check{b}) \quad := \quad \#\Big\{\psi(x_i - t) \leq \check{b}(1-\varepsilon_1),\ U_i = 0\Big\} \tag{8.80}$$

The idea behind Propositions 6.1 and 6.2 is to use the inclusions

$$\Big\{\sum \psi(x_i - t) \leq 0\Big\} \subset \Big\{N_+(t) \leq n_+\Big\},\ \ \Big\{\sum \psi(x_i - t) \geq 0\Big\} \subset \Big\{N_-(t) \leq n_-\Big\} \tag{8.81}$$

for some numbers n_-, n_+ yet to be specified.
For Proposition 6.1, symbolically in the tableau of page 39, we plug in $\delta = 0$, so that the second and third line are separated by $|t| = Cn$. All cases except for case (IV) remain unchanged. For (IV), we consider the first inclusion of (8.81). In this case, $\{\sum \psi(x_i - t) \leq 0\}$ is distorted most importantly by $\tilde{t} = k\hat{b}$. On the other hand the $N'' = n - N_+ - K$ remaining observations cannot be smaller than $N''\check{b}$, so

$$\sum \psi(x_i - t) \leq 0 \qquad \Longrightarrow \qquad N_+\hat{b}(1-\varepsilon_1) + K\hat{b} + N''\check{b} \leq 0 \tag{8.82}$$

that is

$$N_+ \leq \frac{-n\check{b} - K(\hat{b} - \check{b})}{\hat{b}(1-\varepsilon_1) - \check{b}} \tag{8.83}$$

and as this has to hold for all $K \leq \varepsilon_0' n$,

$$N_+ \leq n\frac{-\check{b} - \varepsilon_0'(\hat{b} - \check{b})}{\hat{b}(1-\varepsilon_1) - \check{b}} =: n_+ = n_+(\varepsilon_0') \tag{8.84}$$

where by (6.2) and as $0 < \varepsilon_1 < 1$, $n_+ = n\varepsilon_+$ for

$$0 < \varepsilon_+ = \frac{-\check{b} - \varepsilon_0'(\hat{b} - \check{b})}{\hat{b}(1-\varepsilon_1) - \check{b}} < 1 - \varepsilon_0' \tag{8.85}$$

Accordingly, for the second inclusion in (8.81), we obtain

$$N_- \leq n\varepsilon_- =: n_- = n_-(\varepsilon_0') \qquad \text{for} \quad \varepsilon_- := \frac{\hat{b} - \varepsilon_0'(\hat{b} - \check{b})}{\hat{b} - \check{b}(1-\varepsilon_1)} \tag{8.86}$$

where again $0 < \varepsilon_- < 1 - \varepsilon_0'$. Hence with $\bar{k} = \ulcorner \varepsilon_0' n \urcorner - 1$

$$\begin{aligned} \Pr\{S_n > \sqrt{t}\,\Big|\, D_{k,\tilde{t}=-k\check{b}}\} &\overset{(8.2)}{\leq} \Pr\Big\{T_{n,k}(\sqrt{t}\,) \geq k\check{b}\Big\} \leq \Pr\Big\{T_{n,k}(\sqrt{t}\,) \geq \bar{k}\check{b}\Big\} \leq \\ &\leq \Pr\Big\{N_-(\sqrt{t}\,) \leq n_- \,\Big|\, K = \bar{k}\Big\} \end{aligned} \tag{8.87}$$

and correspondingly

$$\Pr\Big\{S_n < -\sqrt{t}\,\Big|\, D_{k,\tilde{t}=k\hat{b}}\Big\} \leq \Pr\Big\{N_+(-\sqrt{t}\,) \leq n_+ \Big|\, K = \bar{k}\Big\} \tag{8.88}$$

But, $\mathcal{L}(N_{\pm}|K=k)=\mathrm{Bin}(n-k,p_{\pm})$ for

$$p_{-}(t) = \Pr\big(\psi(X^{\rm id}-\sqrt{t}\,)\le \check b(1-\varepsilon_1)\big)=F(\sqrt{t}\,+B_{-}), \qquad (8.89)$$
$$p_{+}(t) = \Pr\big(\psi(X^{\rm id}+\sqrt{t}\,)\ge \hat b(1-\varepsilon_1)\big)=\bar F(-\sqrt{t}\,+B_{+}) \qquad (8.90)$$

where $\bar F=1-F$ and

$$B_{-}:=\inf\Big\{y\,\Big|\,\psi(y)\ge(1-\varepsilon_1)\check b\Big\},\qquad B_{+}:=\sup\Big\{y\,\Big|\,\psi(y)\le(1-\varepsilon_1)\hat b\Big\} \qquad (8.91)$$

If we abbreviate $m=n-\bar k$, $m_{\pm}=\ulcorner n_{\pm}\urcorner$, $p_t=(1-p_{+}(t))\vee p_{-}(t)$, in the binomial probabilities in (8.87)/(8.88), $\binom{m}{j}\le 2^n$, $j=0,\ldots m_{\pm}$, and $p_{-}(t),(1-p_{+}(t))\le 1$, so that

$$\sup_k \Pr\Big\{|S_n|>\sqrt{t}\,\Big|\,D_{k,|\tilde t|=k\check b}\Big\}\le n2^np_t^{[m-(m_-\vee m_+)]} \qquad (8.92)$$

But by (8.85), $1-\varepsilon_0'-(\varepsilon_-\vee\varepsilon_+)=:\alpha>0$, so

$$m-(m_-\vee m_+')\ge \alpha n-1 \qquad (8.93)$$

Now, by (6.4), for $\hat B=\max\{B_+,-B_-\}$, if n is so large that $Cn>(T-\hat B)^2$,

$$\sup_k\int_{Cn}^{\infty}\Pr\Big\{|S_n|>\sqrt{t}\,\Big|\,D_{k,|\tilde t|=kb}\Big\}\le n2^{n+1}\int_{Cn}^{\infty}t^{-\eta(\alpha n-1)/2}\,dt=$$
$$=\exp[-\tilde\alpha n\log(n)(1-\mathrm{o}(n^0))]$$

for some $\tilde\alpha'>0$. So (IV) is indeed negligible. ////

For Proposition 6.2, we only show the first case of (6.5); the second follows analogously. This time $K=0$, n is fixed, and we use the inclusions of the complements in (8.81). Thus

$$\Pr\{S_n\ge\sqrt{t}\}\stackrel{(8.2)}{\ge}\Pr\big\{T_{n,0}(\sqrt{t}\,)>0\big\}\ge\Pr\big\{N_+(\sqrt{t})>n_+(0)\big\}$$

Let $\tilde p_+=\bar F(\sqrt{t}+B_+)$. To $\delta>0$ there is an some $T>0$ such that for $t>T$ and $\tilde p_+^{n_+}>1-\delta$. Hence for $t>T^2$ and $n'=m_++1$

$$\Pr\{S_n>\sqrt{t}\,\}\ge\binom{n}{n'}(1-\tilde p_+)^{n'}\tilde p_+^{n-n'}\ge\binom{n}{n'}(1-\delta)\bar F(\sqrt{t}\,+B_+)^{n'}$$

Now by the first half of (6.5), for $d=1/n'$ and some $c>0$, $T'>T$ and for all $t>T'$

$$t^{1/n'}\big(1-F(t)\big)>c\qquad\Longleftrightarrow\qquad\big(1-F(t)\big)^{n'}>c^{n'}t^{-1} \qquad (8.94)$$

Then for the M-estimator S_n,

$$\mathrm{E}_F[(\mathrm{S}_n)_+^2] \ge \int_{(T')^2}^{\infty}\Pr\Big\{S_n>\sqrt{t}\,\Big\}\,dt$$
$$\ge \int_{(T')^2}^{\infty}\binom{n}{n'}(1-\delta)c^{n'}(\sqrt{t}\,+B_+)^{-1}\,dt=\infty \qquad (8.95)$$

////

8.6 Proof of Proposition 6.3

For $t > v_0^2 \log(n)/n$, we consider the following inclusion

$$\Big\{\psi(x-\sqrt{t}\,) > b - c_0/\sqrt{n}\Big\} = \Big\{x > \sqrt{t} + B_n\Big\} \subset \Big\{x > v_0\sqrt{\log(n)/n} + B_n\Big\}$$

Let

$$A_{k,t} := \Big\{ \sum_{i:\, U_i=1} \psi(X_i - \sqrt{t}\,) \le (k-1)(b - c_0/\sqrt{n}\,)\Big\} \tag{8.96}$$

Hence if $t > v_0^2 \log(n)/n$, by (6.6), for all $k > (1-\delta)r\sqrt{n}$,

$$\Pr(A_{k,t}\,\Big|\,K=k) \ge p_0 \tag{8.97}$$

Now we proceed as in section 8.4, and even with restriction (8.97) the arguments of subsection 8.4.9 remain in force, so that we have to maximize $t^\natural$. But $t > v_0^2\log(n)/n \iff s > \sqrt{\log n}$ in (8.56). Hence on the event $A_{k,t}$ for $s \in [\sqrt{\log n}; bl_n/v_0)$, we get the bound $t^\natural \le (k^\natural - 1)(b - c_0/\sqrt{n})/\sqrt{n}$, while for $s \in (-bl_n/v_0; \sqrt{\log n})$ respectively on ${}^cA_{k,t}$, we bound $t^\natural$ by $k^\natural b$. Integrating out these two s-domains separately as in subsection 8.4.10, we obtain

$$\begin{aligned} & n\Big(\mathrm{MSE}(S_n, Q_n^0\,\big|K=k\,) - \mathrm{MSE}(S_n, Q_n^\flat\,\big|K=k\,)\Big) \ge \\ \ge\quad & p_0\int_{\sqrt{\log n}}^{bl_n/v_0} \Big(2v_0 s D_n(\tilde k) + 2\tilde k b D_n(\tilde k) - D_n(\tilde k)^2\Big)\,\varphi(s)\,ds + \mathrm{o}(n^{-1}) \end{aligned}$$

for

$$D_n(\tilde k) = \tilde k c_0/\sqrt{n} + b/\sqrt{n} + \mathrm{o}(1/\sqrt{n}) \tag{8.98}$$

But for $0 < a_1 < a_2 < \infty$, $\varphi(a_1)/a_2 - \varphi(a_2)/a_2 \le \int_{a_1}^{a_2} \varphi(s)\,ds$, so that with $a_1 = \sqrt{\log n}$, $a_2 = bl_n/v_0$, and as $\varphi(a_2) = \mathrm{o}(n^{-1})$,

$$\begin{aligned} & n\Big(\mathrm{MSE}(S_n, Q_n^0\,\big|K=k\,) - \mathrm{MSE}(S_n, Q_n^\flat\,\big|K=k\,)\Big) \ge \\ \ge\quad & \frac{p_0}{\sqrt{2\pi n}}[2v_0 D_n(\tilde k) - 2\tfrac{\tilde k b D_n(\tilde k) + D_n(\tilde k)^2}{bl_n/v_0}] + \mathrm{o}(n^{-1}) = \frac{2p_0 v_0}{\sqrt{2\pi n}} D_n(\tilde k) + \mathrm{o}(n^{-1}) \end{aligned}$$

Now the restriction to $(1-\delta)r\sqrt{n} < K < k_1 r\sqrt{n}$ by Lemma 8.1 may be dropped, and we obtain

$$n\Big(\mathrm{MSE}(S_n, Q_n^0\,\big|K=k\,) - \mathrm{MSE}(S_n, Q_n^\flat\,\big|K=k\,)\Big) \ge \frac{2p_0 v_0}{n\sqrt{2\pi}}(rc_0 + b) + \mathrm{o}(n^{-1})$$

////

8.7 Proof of Theorem 6.5

In the risk, we have to treat stochastic arguments in Φ, φ; this is settled in the following lemma:

Lemma 8.4 *Let $F\colon \mathbb{R}\to\mathbb{R}$ be twice differentiable with Hölder-continuous second derivative and $G\colon \mathbb{R}\to\mathbb{R}$ be differentiable with Hölder-continuous derivative.*
Then there is a sequence $k_1(n) = 1 + d_n$ with $d_n \to 0$ according to (8.16), such that for all $x, \beta \in \mathbb{R}$ and with $\tilde{k} = K/\sqrt{n}$,

$$\mathrm{E}[F(x+\beta\tilde{k})|K \le k_1(n)r\sqrt{n}\,] = F(x+\beta r) + F''(x+\beta r)\frac{\beta^2 r}{2\sqrt{n}} + \mathrm{o}(n^{-1/2}) \quad (8.99)$$

and

$$\mathrm{E}[G(x+\beta\tilde{k})|K \le k_1(n)r\sqrt{n}\,] = G(x+\beta r) + \mathrm{O}(n^{-(1+\eta)/4}) \quad (8.100)$$

PROOF : Using the Taylor approximation of $\log(1+x)$, we get for n sufficiently large

$$d_n^2/3 \le d_n^2/2 - d_n^3/6 \le \mathcal{K}_n \le d_n^2/2 \quad (8.101)$$

By (8.19) of Lemma 8.1, for some δ_0 and eventually in n we have

$$P(K > k_1(n)r\sqrt{n}\,) \le \exp(-rn^{\delta_0}) \quad (8.102)$$

By the same argument we also get that

$$P(K < (2-k_1(n))r\sqrt{n}\,) \le \exp(-rn^{\delta_0}) \quad (8.103)$$

Hence,

$$P(|\tilde{k} - r| > rd_n) \le 2\exp(-rn^{\delta_0}) \quad (8.104)$$

Thus, as F, G are bounded, the contribution of the set $\{|\tilde{k} - r| > rd_n\}$ decays exponentially, while on the complement we have a uniformly bounded Taylor expansion up to order 2 respectively 1 for the integrands:

$$\begin{aligned} F(x+\beta\tilde{k}) &= F(x+\beta r) + F'(x+\beta r)\beta(\tilde{k}-r) + F''(x+\beta r)\beta^2(\tilde{k}-r)^2/2 + \\ &\quad + \mathrm{o}((\tilde{k}-r)^{2+\eta}) \\ G(x+\beta\tilde{k}) &= G(x+\beta r) + G'(x+\beta r)\beta(\tilde{k}-r) + \mathrm{o}((\tilde{k}-r)^{1+\eta}) \end{aligned}$$

Integrating these expansions out in $\tilde{k}$, we see that the first contribution to the Taylor series for F is the quadratic term, which is $F''(x+\beta r)\frac{\beta^2}{2}\operatorname{Var}\tilde{k}$, and the remainder is $\mathrm{o}(n^{-1/2})$. For G, the first contribution to the error term is the remainder, hence of form $\mathrm{const}|\tilde{k}-r|^{1+\eta}$. By the Hölder inequality this gives a bound

$$\mathrm{const}\,[\operatorname{Var}\tilde{k}]^{\frac{1+\eta}{2}} = \mathrm{O}(n^{-(1+\eta)/4}) \quad (8.105)$$

////

For the proof of Theorem 6.5, we again use the tableau of page 39, albeit with $k_1(n)$ according to (8.16). This time, no integration w.r.t. t is needed, so case (IV) may be cancelled, which is why we may dispense of assumptions (Vb)/(Pd) and pass to the unrestricted neighborhoods $\mathcal{Q}_n$. Cases (II) and (III) may be left unchanged, so we start with working out case (I):

We use α_1, α_2 from (6.18) and proceed paralleling the proof of Theorem 3.6 and get from formula (8.30) that

$$\Pr(S_n \le -\tfrac{\alpha_1}{\sqrt{n}} \,|\, D_{k,\tilde t}\,) = \tilde G_n(\,-\tfrac{\alpha_1}{\sqrt{n}}\,) + \mathrm{O}(n^{-3/2}) \tag{8.106}$$

So we have to spell out $s_{n,k}(\frac{-\alpha_1}{\sqrt{n}}\,)$, which gives

$$s_{n,k}(\tfrac{-\alpha_1}{\sqrt{n}}\,) = v_0^{-1}\Big\{(-t^\natural-\alpha_1)+\tfrac{1}{\sqrt{n}}\big[\tfrac{\tilde k}{2}\alpha_1-\alpha_1\tilde v_1(t^\natural+\alpha_1)-\tfrac{l_2}{2}\alpha_1^2\big]\Big\}+\mathrm{o}(\tfrac{1}{\sqrt{n}}\,) \tag{8.107}$$

and hence —setting $\tilde s = s_{n,k}(\frac{-\alpha_1}{\sqrt{n}}\,)$ and $\tilde s_1 = -(\alpha_1+t^\natural)/v_0$ as in (8.51)

$$\begin{aligned}\Pr(S_n \le -\tfrac{\alpha_1}{\sqrt{n}} \,|\, D_{k,\tilde t}\,) &= \Phi(\tilde s)-\varphi(\tilde s)\tfrac{(\tilde s^2-1)}{6\sqrt{n}}\rho(-\tfrac{\alpha_1}{\sqrt{n}})+\mathrm{o}(\tfrac{1}{\sqrt{n}}\,)=\Phi(\tilde s_1)+\\ &+\tfrac{\varphi(\tilde s_1)}{2\sqrt{n}v_0}[\alpha_1\tilde k-l_2\alpha_1^2-2(\alpha_1+t^\natural)\tilde v_1\alpha_1-v_0\tfrac{\rho_0}{3}(\tilde s_1^2-1)]+\mathrm{o}(\tfrac{1}{\sqrt{n}}\,)\end{aligned} \tag{8.108}$$

This term is maximized eventually in n, if $-t^\natural$ is maximal or, essentially equivalent, all contaminating mass (up to mass $\mathrm{o}(n^{-1/2})$) is concentrated left of $\check y_n$ from (3.3), and then

$$t^\natural = k^\natural \check b \tag{8.109}$$

and after the substitution according to (8.61), this gives with $\tilde s_k = -(\alpha_1+\tilde k\check b)/v_0$

$$\begin{aligned}\Pr(S_n \le -\tfrac{\alpha_1}{\sqrt{n}} \,|\, D_{k,\tilde t=k\check b}\,) &= \Phi(\tilde s_k)+\\ &+\tfrac{\varphi(\tilde s_k)}{2\sqrt{n}v_0}[\alpha_1\tilde k-l_2\alpha_1^2-2\tilde s_k v_0\tilde v_1\alpha_1-v_0\tfrac{\rho_0}{3}(\tilde s_k^2-1)-\tilde k^2\check b]+\mathrm{o}(\tfrac{1}{\sqrt{n}}\,)\end{aligned} \tag{8.110}$$

Now, by (6.22), it holds that $s_1 = -(\alpha_1+r\check b)/v_0$, so that by an application of Lemma 8.4, for $Q^0_{n;\,-}$ any sequence of measures according to (3.23)

$$\begin{aligned}Q^0_{n;\,-}(S_n \le -\tfrac{\alpha_1}{\sqrt{n}}) &= \Phi(s_1)+\mathrm{o}(\tfrac{1}{\sqrt{n}}\,)+\tfrac{1}{\sqrt{n}}\varphi(s_1)\times\\ &\times\big[\tfrac{r}{2v_0}\alpha_1-\tfrac{l_2}{2v_0}\alpha_1^2+s_1v_0\tilde v_1\alpha_1-\tfrac{\rho_0}{6}(\tilde s_1^2-1)-r\tfrac{\check b^2}{2v_0^2}s_1-r^2\tfrac{\check b}{2v_0}\big]\end{aligned} \tag{8.111}$$

Correspondingly, we get for any sequence of measures Q_n^+ according to (3.24)

$$\begin{aligned}Q^0_{n;\,+}(S_n \ge \tfrac{\alpha_2}{\sqrt{n}}) &= \Phi(s_1)+\mathrm{o}(\tfrac{1}{\sqrt{n}}\,)+\tfrac{1}{\sqrt{n}}\varphi(s_1)\times\\ &\times\big[\tfrac{r}{2v_0}\alpha_2+\tfrac{l_2}{2v_0}\alpha_2^2-s_1v_0\tilde v_1\alpha_2+\tfrac{\rho_0}{6}(\tilde s_1^2-1)-r\tfrac{\hat b^2}{2v_0^2}s_1+r^2\tfrac{\hat b}{2v_0}\big]\end{aligned} \tag{8.112}$$

We next account for order $\frac{1}{\sqrt{n}}$-terms and get, as $\delta' = \mathrm{O}(\frac{1}{\sqrt{n}}\,)$

$$Q^0_{n;\,-}(S_n \le -\tfrac{\alpha_1'}{\sqrt{n}}\,) = Q^0_{n;\,-}(S_n \le -\tfrac{\alpha_1}{\sqrt{n}}\,)+\delta'\varphi(\tfrac{a-r\bar b}{v_0})+\mathrm{o}(\tfrac{1}{\sqrt{n}}\,) \tag{8.113}$$

and analogously for $Q^0_{n;\,+}(S_n \ge \frac{\alpha_2'}{\sqrt{n}}\,)$, so

$$\delta' = \tfrac{1}{\sqrt{n}}\Big(-\tfrac{r\delta}{2v_0}-\tfrac{l_2}{2v_0}(a^2+\delta^2)-\tilde v_1v_0s_1\delta-\tfrac{\rho_0}{6}(s_1^2-1)+\tfrac{r\bar b\delta s_1}{v_0^2}+\tfrac{r^2\delta}{2v_0}\Big) \tag{8.114}$$

and $Q^0_{n;-}(S_n \le -\frac{\alpha_1'}{\sqrt{n}}\,) = Q^0_{n;+}(S_n \ge \frac{\alpha_2'}{\sqrt{n}}\,) + \mathrm{o}(\frac{1}{\sqrt{n}}\,)$,

$$\begin{aligned} Q^0_{n;-}(S_n \le -\tfrac{\alpha_1'}{\sqrt{n}}\,) = \Phi(s_1) + \\ +\varphi(s_1)\tfrac{1}{\sqrt{n}}\Big[\tfrac{ra}{2v_0} + 2\tfrac{l_2 a\delta}{v_0} - as_1\tilde v_1 - \tfrac{r(\check b^2+\hat b^2)s_1}{4v_0^2} + \tfrac{r^2\bar b}{2v_0}\Big] + \mathrm{o}(\tfrac{1}{\sqrt{n}}\,) \end{aligned} \tag{8.115}$$

////

8.8 Proof of Corollary 7.2

The assumptions of Theorem 7.1 are clearly fullfilled. Hence we may start with the verification (7.4):

$$G(w,s) = (w^2+s^2)(1+\frac{r}{\sqrt{n}}\,) + \frac{r}{\sqrt{n}}\,w^2(1+\frac{1}{r^2}) \tag{8.116}$$

$$\partial_w G(w,s) = 2w[1+\frac{r}{\sqrt{n}} + \frac{r}{\sqrt{n}}(1+\frac{1}{r^2})\,] \tag{8.117}$$

$$\partial_s G(w,s) = 2s[1+\frac{r}{\sqrt{n}}\,] \tag{8.118}$$

and hence, dividing both sides of (7.3) by $2\hat A\hat v_0$, we get the assertion. The LHS of (7.4) (with or without factor $1+\frac{r^2+1}{r^2+r\sqrt{n}}$) is isotone, the RHS antitone in c. Thus if we insert the factor to correct the f-o-o clipping height c_0 to $c_1(n)$, the factor increases the LHS without affecting the RHS. This can only be compensated for by a decrease of c_0 to $c_1(n)$. If $h(c)$ is differentiable in c_0 with derivative $h'(c_0)$, (7.5) is an application of the applying the implicit function theorem: Let $G(s,c) := r^2c\,(1+s) - h(c)$. Then $G(0,c_0)=0$. Hence for $s=\frac{r^2+1}{r^2+r\sqrt{n}}$, up to $\mathrm{o}(n^{-1/2})$,

$$c_1(n) + \mathrm{o}(n^{-1/2}) = c_0 - \frac{G_s(0,c_0)}{G_c(0,c_0)}s = c_0\left(1-\frac{1}{\sqrt{n}}\,\frac{r^3+r}{r^2-h'(c_0)}\right) + \mathrm{o}(n^{-1/2})$$

////

8.9 Proof of Proposition 7.3

We apply Rieder (1994, Theorem 1.4.7) to the derivatives; this theorem says that for $\eta\in\mathcal{C}_1(\mathbb{R})$ with $\eta(\theta_0)=0$, $\eta'(\theta_0)\ne 0$ for some $\theta_0\in\mathbb{R}$, there exists an open neighborhood $V_0\subset\mathcal{C}_1(\mathbb{R})$ such that for every open, connected neighborhood $V\subset V_0$ of η there is a unique, continuous map $T\colon V\to\mathbb{R}$ with

$$T(\eta)=\theta_0,\qquad f(T(f))=0,\quad f\in V \tag{8.119}$$

even more so, T is continuously bounded differentiable on V with derivative at tangent h

$$dT(f)h = -h(T(f))/f'(T(f)) \tag{8.120}$$

Hence there is an open neighborhood $V_{0;F}$ of F such that for each connected open neighborhood $V_F \subset V_{0;F}$, we get a unique, continuously bounded differentiable map $T\colon V_F \to \mathbb{R}$ with

$$T(F)=x_0,\quad f'(T(f))=0,\quad f\in V_F,\quad dT(f)h=-h'(T(f))/f''(T(f)) \tag{8.121}$$

But by assumption (7.6) from some n on, F_n and G_n will lie in $V_{0;F}$, and setting $x_n = T(F_n)$, by (8.121) $F_n'(x_n)=0$, and

$$|x_n - x_0| = |T(F_n)-T(F)| \le |F_n'(x_0)|/F''(x_0) = \mathrm{O}(n^{-\beta'})$$

which is (b); again by (7.6),

$$\begin{aligned} |F_n''(x_n)-F''(x_0)| &\le |F_n''(x_n)-F''(x_n)|+|F''(x_n)-F''(x_0)| \le \\ &\le \sup_x |F_n''(x)-F''(x)| + \mathrm{o}(n^0) = \mathrm{o}(n^0) \end{aligned}$$

In particular, eventually in n, $F_n''(x_n)>0$ and hence x_n is a minimum of F, so (a) is shown. By (7.6), $\sup_x |F-G_n|+|F'-G_n'|+|F''-G_n''| = \mathrm{O}(n^{-\beta'})$, so (c) follows just as (a). For (d) we note

$$|x_n-y_n| = |T(F_n)-T(G_n)| \le |G_n'(x_n)|/F_n''(x_n) \stackrel{(a)}{=} |G_n'(x_n)|/(f_2+\mathrm{o}(n^0)) = \mathrm{O}(n^{-\beta})$$

To show (e), we introduce $d_n := y_n - x_n$ and write

$$\begin{aligned} 0 &\le G_n(x_n)-G_n(y_n) = G_n'(y_n)d_n + G_n''(y_n)d_n^2/2 + \mathrm{o}(d_n^2) = \\ &= (f_2+\mathrm{o}(n^0))d_n^2/2 + \mathrm{o}(d_n^2) = \mathrm{O}(n^{-2\beta}) \end{aligned} \tag{8.122}$$

////

8.10 Proof of Proposition 7.8

We show that under the assumptions of this proposition x_j indeed defines a "uniformly bad contamination" in the sense that for the fixed contamination $Q_n(x_j)$

$$\mathrm{asMSE}_0(S_n^{(b_0)}, Q_n(x_0)) = \min_{b>0} \mathrm{asMSE}_0(S_n^{(b)}, Q_n(x_0))) \tag{8.123}$$

resp.

$$\mathrm{asMSE}_1(S_n^{(c_1)}, Q_n(x_1)) = \min_{c>0} \mathrm{asMSE}_1(S_n^{(c)}, Q_n(x_1))) \tag{8.124}$$

In case $j=0$, as in the setup of Rieder (1994, chap. 5), we obtain

$$\mathrm{asMSE}_0(S_n^{(b)}, Q_n(x_0)) = \operatorname{tr}\mathrm{Cov}_{\mathrm{id}}(\eta_b) + r^2|\eta_b(x_0)|^2 \tag{8.125}$$

and

$$\mathrm{asMSE}_0(\hat S_n, Q_n(x_0)) = \operatorname{tr}\mathcal{I} + r^2|\hat\eta(x_0)|^2 \tag{8.126}$$

Now for given x_0, either $|\eta^{(b)}(x_0)|<b$ or $|\eta^{(b)}(x_0)|=b$. In the first case, (7.23) applies and hence

$$\mathrm{asMSE}_0(S_n^{(b_0)}, Q_n(x_0)) \ge \mathrm{asMSE}_0(\hat S_n, Q_n(x_0)) \tag{8.127}$$

In the latter, $Q_n(x_0)$ already achieves maximal asymptotic risk for $S_n^{(b)}$ on $\mathcal{Q}_n$, and hence by minimaxity of $S_n^{(b_0)}$

$$\mathrm{asMSE}_0(S_n^{(b)}, Q_n(x_0)) \geq \mathrm{asMSE}_0(S_n^{(b_0)}, Q_n(x_0)) \tag{8.128}$$

For the case $j = 1$ one argues in an analogue way. ////

8.11 Proof for (7.23) in the Gaussian location scale model

We abbreviate the location and scale parts by indices l and s respectively. By equivariance we may limit ourselves to the case $\theta = (0,1)^\tau$. Due to symmetry, $A = A(b)$ from (1.9) is diagonal for all b with elements A_l and A_s and we may write

$$\eta_b = Y \min\{1, b/|Y|\}, \qquad Y^\tau = \big(A_l x, A_s(x^2 - 1 - z_s)\big) \tag{8.129}$$

The centering $z_s(b)$ after the clipping is necessary, as the scale part is not skew symmetric; in the pure scale case (with known θ_l), the corresponding centering $z_s' = z_s'(b)$ is antitone in b, because Λ_s is monotone in x^2: It decreases from 0 to $[\Phi^{-1}(3/4)]^2 - 1 \doteq -0.545 =: \check{z}$. In the combined case, we never reach this extremal case due to the additional location part —compare Kohl (2005, Remark 8.2.1(a)) where $\bar{z}_s = \bar{a}_{\mathrm{sc}}/\bar{\alpha} - 1 \doteq -0.530$; in any case, $z_s > -1$ always. Hence in particular, for $x_0 = 1.844$ and b such that $|\eta^{(b)}(x_0)| \leq b$ it holds that

$$|\eta_s^{(b)}(x_0)| = A_s(b)|x_0^2 - 1 - z_s(b)| > A_s(b)|x_0^2 - 1| > \mathcal{I}_s^{-1}|x_0^2 - 1| = |\hat{\eta}_s(x_0)| \tag{8.130}$$

and thus in particular,

$$\begin{aligned} |\eta^{(b)}(x_0)|^2 &= |\eta_s^{(b)}(x_0)|^2 + |\eta_l^{(b)}(x_0)|^2) = |\eta_s^{(b)}(x_0)|^2 + A_{0;l}(b)x_0^2 > \\ &> \hat{\eta}_s(x_0)^2 + \mathcal{I}_l^{-2}x_0^2 = |\hat{\eta}(x_0)|^2 \end{aligned} \tag{8.131}$$

////

9 Appendix

9.1 Two Hoeffding Bounds

Lemma 9.1 *Let* $\xi_i \stackrel{\text{i.i.d.}}{\sim} F$, $i = 1, \ldots, n$ *be real–valued random variables,* $|\xi_i| \leq M$ *Then for* $\varepsilon > 0$

$$P(\frac{1}{n}\sum_i \xi_i - \mathrm{E}[\xi_1] \geq \varepsilon) \quad \leq \quad \exp(-\frac{2n\varepsilon^2}{M^2}) \tag{9.1}$$

$$P(\frac{1}{n}\sum_i \xi_i - \mathrm{E}[\xi_1] \leq -\varepsilon) \quad \leq \quad \exp(-\frac{2n\varepsilon^2}{M^2}) \tag{9.2}$$

Proof : Hoeffding (1963), Thm. 2. ////

Lemma 9.2 *Let* $\xi_i \overset{\text{i.i.d.}}{\sim} F$, $i = 1, \ldots, n$ *be real–valued random variables,* $|\xi_i| \le 1$ *Then for* $\mu = \mathrm{E}[\xi_1]$ *and* $0 < \varepsilon < 1 - \mu$

$$P(\frac{1}{n}\sum_i \xi_i - \mu \ge \varepsilon) \quad \le \quad \left\{ \left(\frac{\mu}{\mu+\varepsilon}\right)^{\mu+\varepsilon} \left(\frac{1-\mu}{1-\mu-\varepsilon}\right)^{1-\mu-\varepsilon} \right\}^n \tag{9.3}$$

PROOF : Hoeffding (1963), Thm. 1, inequality (2.1). ////

9.2 A uniform Edgeworth expansion

In the following theorem, we generalize Ibragimov (1967, Thm. 1) and Ibragimov and Linnik (1971, Thm. 3.3.1) to the situation where the law of ξ_i depends through an additional parameter t:

Theorem 9.3 *For some set* $\Theta \subset \mathbb{R}$ *and fixed* $t \in \Theta$ *let* $\xi_{i,t}$, $i = 1, 2, \ldots$ *be a sequence of i.i.d. real-valued random variables with distribution* F_t *and with*

$$\mathrm{E}\,\xi_{i,t} = 0, \quad \mathrm{E}\,\xi_{i,t}^2 = 1, \quad \mathrm{E}\,\xi_{i,t}^3 = \rho_t, \quad \mathrm{E}\,\xi_{i,t}^4 - 3 = \kappa_t \tag{9.4}$$

Let $\Phi(s)$ *and* $\varphi(s)$ *be the c.d.f. and p.d.f. of* $\mathcal{N}(0,1)$ *and*

$$F_n(s,t) \quad := \quad P(\textstyle\sum_{i=1}^n \xi_{i,t} < s\sqrt{n}) \tag{9.5}$$

$$H_n(s,t) \quad := \quad \Phi(s) - \frac{\varphi(s)}{\sqrt{n}}\varphi(s)\frac{\rho_t}{6}(s^2-1) \tag{9.6}$$

$$G_n(s,t) \quad := \quad H_n(s,t) - \frac{\varphi(s)}{n}\Big[\frac{\kappa_t}{24}(s^3-3s) + \frac{\rho_t^2}{72}(s^5-10s^3+15s)\Big] \tag{9.7}$$

Let f_t *be the characteristic function of* F_t.

(a) *If* $\sup_t \kappa_t < \infty$ *and if there is some* $u_0 > 0$ *such that for all* u_1 *the "no-lattice"-condition (C)'*

$$\hat{f}_{u_0}(u_1) := \sup_{u_0<u<u_1} \sup_t |f_t(u)| < 1 \tag{9.8}$$

is fulfilled, then

$$\sup_{s\in\mathbb{R}} \sup_t |F_n(s,t) - H_n(s,t)| = \mathrm{o}(n^{-1/2}) \tag{9.9}$$

(b) *If*

$$\sup_t \mathrm{E}\,|\xi_{i,t}|^5 < \infty \tag{9.10}$$

and the uniform Cramér–condition (C)

$$\limsup_{u\to\infty} \sup_t |f_t(u)| < 1 \tag{9.11}$$

is fulfilled, then

$$\sup_{s\in\mathbb{R}} \sup_t |F_n(s,t) - G_n(s,t)| = \mathrm{O}(n^{-3/2}) \tag{9.12}$$

PROOF : The general technique to prove Edgeworth expansions is to use Berry's smoothing lemma, which we take from Ibragimov and Linnik (1971, Thm. 1.5.2): and apply it to our case: Let $f_{n,t}$ be the characteristic function of $F_n(\,\cdot\,,t)$, and define the Edgeworth measures $G_{n,j,t}$, $j=1,2$ as

$$G_{n,2,t}(s) = H_n(s,t), \qquad G_{n,3,t}(s) = G_n(s,t)$$

as well as their Fourier-Stieltjes transforms

$$g_{n,j,t}(u) = \int e^{isu} G'_{n,j,t}(s)\,\lambda(ds) \tag{9.13}$$

and

$$\hat G'_{n,j} = \sup_t \sup_{s\in\mathbb{R}} |G'_{n,j,t}(s)| \tag{9.14}$$

Then for $T > T' > 0$, it holds that

$$\begin{aligned}
&\sup_{s\in\mathbb{R}}\sup_t |F_n(s,t) - G_{n,j,t}(s)| \le \\
\le\quad & \sup_t \frac{1}{\pi}\int_{-T'}^{T'} \frac{|f_{n,t}(u) - g_{n,j,t}(u)|}{|u|}\,\lambda(du) + \sup_t \frac{1}{\pi}\int_{T'\le |u| < T} \frac{|f_{n,t}(u)|}{|u|}\,\lambda(du) + \\
& + \sup_t \frac{1}{\pi}\int_{T'\le |u| < T} \frac{|g_{n,j,t}(u)|}{|u|}\,\lambda(du) + \sup_t \frac{24}{\pi T}\hat G'_{n,j}
\end{aligned} \tag{9.15}$$

But similarly as in Ibragimov (1967, p. 462/3), for some constants $\gamma > 0$ and $c_j > 0$, we get for $T' = \gamma\sqrt{n}$ and $|u| \le T'$

$$\frac{|f_{n,t}(u) - g_{n,j,t}(u)|}{|u|} \le c_j \sup_t \mathrm{E}[|\xi_{1,t}|^{3+j}]\, n^{-(j+1)/2}\,(|u|^j + |u|^{2+3j})\, e^{-u^2/4} \tag{9.16}$$

and hence the first summand in the RHS of (9.15) is $\mathrm{O}(n^{-(j+1)/2})$. For the second summand, we note that $f_{n,t}(u) = f_t^n(u/\sqrt{n})$ and hence

$$\int_{T'}^{T} \frac{|f_{n,t}(u)|}{u}\,\lambda(du) = \int_{\gamma}^{T/\sqrt{n}} \frac{|f_t^n(u)|}{u}\,\lambda(du) \tag{9.17}$$

In case $j=2$, for γ sufficiently large, by condition (C), $\sup_t \sup_{|u|>\gamma} |f_t(u)| =: \beta < 1$ and hence, for $T = n^{3/2}$,

$$\sup_t \int_{T'}^{T} \frac{|f_{n,t}(u)|}{u}\,\lambda(du) \le \log(T/\sqrt{n}\,)\beta^n = \mathrm{o}(e^{-\sqrt{n}/2}) \tag{9.18}$$

In case $j=1$, we proceed as in Ibragimov and Linnik (1971, Lemma 3.3.1): If $\sup_{u_1} \hat f_\gamma(u_1) < \infty$ for γ sufficiently large, we may proceed as in case $j=2$; else, (C') says that for γ sufficiently large, $\hat f_\gamma(u_1)$ is isotone and tends to 1. So we may define

$$l'_n := \inf\{u_1 \,\big|\, \hat f_\gamma(u_1) \ge 1 - 1/\sqrt{n}\,\} \tag{9.19}$$

Setting $T = \sqrt{n}\, l_n$ for $l_n = \min(l'_n, \sqrt{n})$, we see that $l_n^{-1} = \mathrm{o}(n^0)$ and

$$\sup_t \int_{T'}^{T} \frac{|f_{n,t}(u)|}{u}\,\lambda(du) \le \log(\sqrt{n})(1-1/\sqrt{n})^n \le \log(\sqrt{n})e^{-\sqrt{n}} = \mathrm{o}(e^{-\sqrt{n}/2}) \tag{9.20}$$

Hence the second summand in the the RHS of (9.15) is $\mathrm{O}(n^{-(j+1)/2})$. Also, it is easy to see that $\hat G'_{n,j} < \infty$, and hence by the choice of T, the last summand in the the RHS of (9.15) is $\mathrm{O}(l_n^{-1} n^{-1/2}) = \mathrm{o}(n^{-1/2})$ in case $j=1$, and $\mathrm{O}(n^{-3/2})$ for $j=2$. Finally, by Mill's ratio, the third summand is again easily shown to be $\mathrm{O}(\exp(-\gamma^2 n/3))$. ////

9.3 Moments for the Binomial

Lemma 9.4 *Let $X \sim \mathrm{Bin}(n,p)$. Then*

$$\mathrm{E}[X] = pn, \qquad \mathrm{E}[X^2] = p^2n^2 + pn - p^2n, \tag{9.21}$$

$$\mathrm{E}[X^3] = p^3n^3 - 3p^3n^2 + 2p^3n + 3p^2n^2 - 3p^2n + pn, \tag{9.22}$$

$$\begin{aligned}\mathrm{E}[X^4] = {} & p^4n^4 - 6p^4n^3 + 11p^4n^2 - 6p^4n + \\ & + 6p^3n^3 - 18p^3n^2 + 12p^3n + 7p^2n^2 - 7p^2n + pn\end{aligned} \tag{9.23}$$

and consequentially, for $p = r/\sqrt{n}$,

$$\mathrm{E}[X] = rn^{1/2}, \qquad \mathrm{E}[X^2] = r^2n + rn^{1/2} - r^2, \tag{9.24}$$

$$\mathrm{E}[X^3] = r^3n^{3/2} + 3r^2n + (r - 3r^3)n^{1/2} - 3r^2 + 2r^3n^{-1/2}, \tag{9.25}$$

$$\begin{aligned}\mathrm{E}[X^4] = {} & r^4n^2 + 6r^3n^{3/2} + (7r^2 - 6r^4)n + (r - 18r^3)n^{1/2} + \\ & + 11r^4 - 7r^2 + 12r^3n^{-1/2} - 6r^4n^{-1}\end{aligned} \tag{9.26}$$

PROOF : easy calculations for `MAPLE` — see procedure `Binmoment`... ////

9.4 Decay of the standard normal

Finally, we note the following Lemma for $\mathcal{N}(0,1)$ variables

Lemma 9.5 *Let $X \sim \mathcal{N}(0,1)$. Then for $k = 0,1,2,\ldots,8$ and any sequence $(c_n)_n \subset \mathbb{R}$ with $\liminf_n c_n > \sqrt{2}$,*

$$\mathrm{E}[X^k\, \mathrm{I}_{\{X \ge c_n\sqrt{\log(n)}\}}] = \mathrm{o}(n^{-1}) \tag{9.27}$$

PROOF : Let $\Phi(x) := \Pr(X \le x)$, $\bar\Phi(x) := \Pr(X > x)$, $\varphi(x)$ the density of X. Then

$$\mathrm{E}[X^k\, \mathrm{I}_{\{X \ge 2\sqrt{\log(n)}\}}] = \left\{\begin{array}{ll} P_k(x)\,\varphi(x)\Big|_{c\sqrt{\log(n)}}^{\infty} & \text{for } \; k \text{ odd} \\ P_k(x)\,\varphi(x) + \prod_{i=1}^{k/2}(2i-1)\Phi(x)\Big|_{c\sqrt{\log(n)}}^{\infty} & \qquad k \text{ even} \end{array}\right.$$

for some polynomial P_k of degree $k-1$. The assertion follows, as

$$\varphi(c\sqrt{\log(n)}) = \varphi(0)n^{-c^2/2} = \varphi(0)n^{-(1+\delta)}$$

for some $\delta > 0$, and because for the $\Phi(x)$-term Mill's ratio applies: That is, $\bar{\Phi}(x) \le \varphi(x)/x$ for $x > 0$. ////

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, N. J. 5.3.1, 7.4

Barndorff-Nielsen O. and Cox D. (1994): *Inference and asymptotics.*, Vol. 52 of *Monographs on Statistics and Applied Probability*. Chapman and Hall. 4.1

Bhattacharya R. and Rao R. (1976): *Normal approximation and asymptotic expansions.*. Wiley Series in Probability and Mathematical Statistics. John Wiley&Sons, Inc. 4.1

Bickel P.J., Klaassen C.A., Ritov Y. and Wellner J.A. (1998): *Efficient and adaptive estimation for semiparametric models.*. Springer. 1.4

Daniels H.E. (1954): Saddlepoint approximations in statistics. *Ann. Math. Statistics*, **25**: 631–650. 4.1

Donoho D.L. and Huber P.J. (1983): The notion of breakdown point. In: *A Festschrift for Erich L. Lehmann*, (P.J. Bickel, K. Doksum and J.L. Hodges, Jr., eds.), p. 157–184. Wadsworth, Belmont, CA. (document), 1.6, 2.2

Field C. and Ronchetti E. (1990): *Small sample asymptotics*, Vol. 13 of *IMS Lecture Notes - Monograph Series.*. Institute of Mathematical Statistics., Hayward, CA. 4.1, 4.1

Fraiman R., Yohai V.J. and Zamar R.H. (2001): Optimal robust M-estimates of location. *Ann. Stat.*, **29**(1): 194–223. 1.6, 3.5.1, 4.2, 6.3

Gusev, S.I. (1976): Asymptotic expansions associated with some statistical estimators in the smooth case. II: Expansions of moments and distributions. *Theor. Probability Appl.*, **21**: 14–33. 4.1

Hall P. (1992): *The bootstrap and Edgeworth expansion.*. Springer Series in Statistics. Springer-Verlag. 1.5, 4.1

Hampel F.R. (1974): Some small sample asymptotics. In: *Proc. Prague Symp. Asympt. Stat., Vol. II, Prague 1973*, p. 109–126. 4.1

Hoeffding W. (1963): Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.*, **58**: 13–30. 9.1, 9.1

Huber P.J. (1968): Robust confidence limits. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **10**: 269–278. (document), 1.6, 6.3

—— (1981): *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley. 1.5, 2.2, 8.1

—— (1997): *Robust statistical procedures.*, Vol. 68 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2. edition. (document), 7.5

Ibragimov I. (1967): the Chebyshev-Cramér asymptotic expansions. *Theor. Probab. Appl.*, **12**: 454–469. 9.2, 9.2

Ibragimov I. and Linnik Y. (1971): *Independent and stationary sequences of random variables*. Wolters-Noordhoff Series of Monographs and Textbooks on Pure and Applied Mathematics. Wolters-Noordhoff Publishing Company., Groningen. Edited by J.F.C. Kingman. 4.1, 9.2, 9.2, 9.2

Kohl M. (2005): *Numerical contributions to the asymptotic theory of robustness*. Dissertation, Universität Bayreuth, Bayreuth. b, 6.4, 7.5.4, 8.11

Kohl M., Ruckdeschel P. and Stabla T. (2004): General Purpose Convolution Algorithm for Distributions in S4-Classes by means of FFT. Unpublished manuscript. Also available in
`http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/comp.pdf.`
1.6, 5.2

Le Cam L. (1986): *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer. 1.4

Pfanzagl J. (1979): First order efficiency implies second order efficiency. In: *Contributions to statistics, Jaroslav Hajek Mem. Vol.*, p. 167-196. 1.6

—— (1985): *Asymptotic expansions for general statistical models. With the assist. of W. Wefelmeyer.*, Vol. 31 of *Lecture Notes in Statistics*. Springer-Verlag. 4.1

Pfaff T. (1977): *Existenz und asymptotische Entwicklungen der Momente mehrdimensionaler maximum likelihood-Schätzer*, Dissertation, Universität zu Köln, Köln. 4.1

R Development Core Team (2005): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org* 7.3.2

Rieder H. (1980): Estimates derived from robust tests. *Ann. Stat.*, **8**: 106–115. (document), 1.6, 6.3, 6.3, 6.3

—— (1994): *Robust asymptotic statistics*. Springer Series in Statistics. Springer. (document), 1.1, 1.2, 1.3, 1.4, 2, 6.4, 8.1, 8.9, 8.10

Rieder H., Kohl M. and Ruckdeschel P. (2001): The Costs of not Knowing the Radius. Submitted. Appeared as discussion paper Nr. 81. SFB 373 (Quantification and Simulation of Economic Processes), Humboldt University, Berlin; also available in
`http://www.uni-bayreuth.de/departments/math/org/mathe7/RIEDER/pubs/RR.pdf`. (document), 1.6, 7.4

Ruckdeschel P. (2004): A Motivation for $1/\sqrt{n}$–Shrinking–Neighborhoods. To appear in Metrika. Also available in
`http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/whysqn.pdf`. 7.5, 7.5.1

—— (2005a): Higher order asymptotics for the MSE of the median on shrinking neighborhoods. Unpublished manuscript. Also available in
`http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/medmse.pdf`. b, b, d, 3.5.2, 3.5.2, 3.5.3, 5.3.3, 8.4.5

—— (2005b): Higher order asymptotics for the MSE of one-step estimators on shrinking neighborhoods. Unpublished manuscript. 1.7, 7.5.2

Ruckdeschel P. and Kohl M. (2004): How to approximate finite sample risk of M-Estimators. Unpublished manuscript. Also available in
`http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/howtoap.pdf`. 4.1, 4.2, 5.2

Ruckdeschel P. and Rieder, H. (2004): Optimal Influence Curves for General Loss Functions. Statistics and Decisions. **22**: 201–223. 7.1

Ruckdeschel P., Kohl M., Stabla T. and Camphausen F. (2004): `S4` *Classes for Distributions*. Submitted. Also available in
`http://www.uni-bayreuth.de/departments/math/org/mathe7/DISTR`. 5.2

Taniguchi M. and Kakizawa Y. (2000): *Asymptotic theory of statistical inference for time series.*. Springer Series in Statistics. Springer. 4.1

van der Vaart A. (1998): *Asymptotic statistics.*, Vol. 3 of *Cambridge Series on Statistical and Probabilistic Mathematics*. Cambridge Univ. Press., Cambridge. 1.4

Web-page to this article:
`http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/mest.html`