Higher Order Expansion for the MSE of M-estimators on shrinking neighborhoods

Peter Ruckdeschel

Received: date / Accepted: date

Abstract We consider estimation of a one-dim. location parameter by means of M-estimators S_n with monotone influence curve ψ. For growing sample size n, on suitably thinned out convex contamination balls $\tilde{Q}_n$ of shrinking radius $r/\sqrt{n}$ about the ideal distribution, we obtain an expansion of the asymptotic maximal mean squared error MSE of form

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

where A_1, A_2 are constants depending on ψ and r. Hence S_n not only is uniformly (square) integrable in n (in the ideal model) but also on $\tilde{Q}_n$, which is not self-evident. For this result, the thinning of the neighborhoods, by a breakdown-driven, sample-wise restriction, is crucial, but exponentially negligible. Moreover, our results 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.

Keywords higher order asymptotics · location M-estimator · uniform integrability · Edgeworth expansion · gross error neighborhood · shrinking neighborhood · breakdown point

Mathematics Subject Classification (2000) MSC 62F35, 62F12

1 Motivation/introduction

In the setup of shrinking neighborhoods about a general, parametric ideal central model, Rieder [22] determines the asymptotically linear estimator (ALE) minimax-

P. Ruckdeschel
Fraunhofer ITWM, Department of Financial Mathematics,
Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
and Dept. of Mathematics, University of Kaiserslautern,
P.O.Box 3049, 67653 Kaiserslautern, Germany
E-mail: peter.ruckdeschel@itwm.fraunhofer.de

ing as. MSE on these neighborhoods. We address the question to which degree this asymptotic optimality carries over to finite sample size and try to identify and quantify which aspects of both estimator and neighborhood are responsible for the quality of the approximation.

1.1 Setup: one-dimensional location

As a starting point for assessing such questions we consider the most basic parametric model of statistics, the one-dimensional location model $\{P_\theta(dx) = F(dx - \theta),\ \theta \in \mathbb{R}\}$ for some ideal distribution F with finite Fisher-Information of location $\mathcal{I}(F)$ in the sense of Huber [14, 4.Def.4.1, Thm.4.2], i.e. $\mathcal{I}(F) := \sup_{\varphi \in C_c^1} (\int \dot\varphi\, dF)^2 / (\int \varphi^2\, dF)$, entailing that $\Lambda_f = -\dot f/f \in L_2(F)$, $\mathcal{I}(F) = \mathrm{E}[\Lambda_f^2]$. Paralleling Huber [13], we also assume that Λ_f is increasing. By translation equivariance, we may restrict ourselves to $\theta_0 = 0$ which is suppressed in the notation.

The set of *influence curves* (IC's) Ψ in this model is defined as in Rieder [22]

$$\Psi := \{\psi \in L_2(F) \,|\, \mathrm{E}[\psi] = 0, \quad \mathrm{E}[\psi \Lambda_f] = 1\}, \tag{1.1}$$

where both expectations are evaluated under F.

Shrinking neighborhoods Robust Statistics enlarges the ideal model assumptions by suitable neighborhoods about them. The shrinking neighborhood approach—compare e.g. Rieder [22], Kohl et al. [17], balances bias and variance, which would be of different scaling in n otherwise, see also Ruckdeschel [24]. For this paper we consider contamination neighborhoods, i.e. the set $\mathcal{Q}_n(r)$ of distributions

$$\mathcal{L}_\theta^{\mathrm{real}}(X_1, \ldots, X_n) = Q_n = \bigotimes_{i=1}^n [(1 - \tfrac{r_n}{\sqrt{n}})F + \tfrac{r_n}{\sqrt{n}}\, P_{n,i}^{\mathrm{di}}] \tag{1.2}$$

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

$$X_i := (1 - U_i)X_i^{\mathrm{id}} + U_i X_i^{\mathrm{di}}, \qquad i = 1, \ldots, n \tag{1.3}$$

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

First order optimality For a sequence of estimators S_n, consider as risk the asymptotically (modified) maximal MSE on $\mathcal{Q}_n$

$$\tilde R(S_n, r) := \lim_{t\to\infty} \lim_{n\to\infty} \sup_{Q_n \in \mathcal{Q}_n(r)} \int \min\{t,\, n\,|S_n - \theta_0|^2\}\, dQ_n \tag{1.4}$$

Following Rieder [22, Ch. 5] a (suitably constructed) ALE S_n with IC ψ has risk

$$\tilde R(S_n, r) = r^2 \sup |\psi|^2 + \mathrm{E}_{\mathrm{id}} |\psi|^2 \tag{1.5}$$

By Theorem 5.5.7 (ibid.), together with its preceding remarks, for given $r \geq 0$, a (suitably constructed) ALE with IC $\hat{\eta}$ minimizes $\tilde{R}(\,\cdot\,, r)$ among all ALE's iff $\hat{\eta} = \eta_{c_0}$ for Lagrange multipliers z and A such that η_{c_0} is an IC for

$$\eta_{c_0} = A(\Lambda_f - z)\min\{1, c_0/|\Lambda_f - z|\}, \qquad \mathrm{E}[(|\Lambda_f - z| - c_0)_+] = r^2 c_0 \tag{1.6}$$

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 (1.4) of the MSE, which is common in asymptotic statistics, cf. Le Cam [18], Rieder [22], Bickel et al. [4], van der Vaart [31], and which forces the integrals to converge under weak convergence, has no statistical justification. One would perhaps prefer a modification that is statistically motivated.

1.2 M-estimators for location

There are several constructions for an ALE to achieve a given IC ψ—one-step constructions, M-estimators, L-estimators and many more. In this paper we confine ourselves to M-estimators. We require ψ to be monotone and bounded and write $\psi_t(\,\cdot\,)$ for $\psi(\,\cdot\, - t)$. For technical reasons we assume that the set D_t of discontinuities of the c.d.f. of $\psi_t(X^{\mathrm{id}})$ has to carry less mass than 1 uniformly:

$$p_D := \sup\nolimits_t P^{\mathrm{id}}(D_t) < 1 \tag{1.7}$$

Following the notation in Huber [14, pp. 46], let

$$S_n^{*} := \sup\Big\{t \,|\, \sum_{i\leq n}\psi_t(x_i) > 0\Big\}, \qquad S_n^{**} := \inf\Big\{t \,|\, \sum_{i\leq n}\psi_t(x_i) < 0\Big\} \tag{1.8}$$

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

$$\Pr\{S_n^{*} < t\} = \Pr\Big\{\sum_{i\leq n}\psi_t(x_i) \leq 0\Big\}, \qquad \Pr\{S_n^{**} < t\} = \Pr\Big\{\sum_{i\leq n}\psi_t(x_i) < 0\Big\} \tag{1.9}$$

in the continuity points t of the LHS. The next lemma, an immediate consequence of Hall [10, Theorem 2.3], shows that we may ignore the event $S_n^{*} \neq S_n^{**}$ if we are interested in statements valid up to $\mathrm{o}(1/n)$.

Lemma 1.1 *Under* (1.7), $\Pr(S_n^{*} \neq S_n^{**}) = \mathrm{O}(\exp(-\gamma n))$ *for some* $\gamma > 0$.

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

1.3 Organization of this paper and description of the results

This paper provides answers to some of the open questions mentioned in subsection 1.1; these answers were initiated by an attempt to check the validity of Rieder's asymptotic approach at finite sample sizes by simulations in 2003. At closer inspection of these simulations, M. Kohl found out that larger inaccuracies of (first order) asymptotics only occurred in 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. With regard to our shrinking setup, such an exclusion on the one hand is asymptotically negligible, hence does not affect the results of subsection 1.1, but on the other hand under this restriction indeed the unmodified MSE converges along with weak convergence. We discuss this modification in section 2. In section 3, we present the central theoretical result, Theorem 3.5. This result is of the following form

$$\sup_{Q_n\in\tilde{Q}_n(r;\varepsilon_0)} n\,\mathrm{MSE}(S_n,Q_n) = r^2\sup|\psi|^2 + \mathrm{E}\,\psi^2 + \frac{r}{\sqrt{n}}\,A_1 + \frac{1}{n}\,A_2 + \mathrm{o}(\tfrac{1}{n}) \qquad (1.10)$$

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 at once that the speed of the convergence to the first order asymptotic value is one order faster in the ideal model.

Notation 1.3 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. The relevance of these results for (small) finite sample sizes is shown by a simulation study which is presented in section 5 as to its design and results. By means of an adopted convolution algorithm taken from Ruckdeschel and Kohl [27], we also compute numerically exact values of the MSE. Proofs are delegated to the appendix section A. 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 B.

On a web-page to 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.

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. LAN holds for L_2-differentiable models, cf. Rieder [22, Thm. 2.3.5]. and together with LeCam's third Lemma—cf. Cor. 2.2.6 ibid.—implies

[1] for local asymptotic normality

uniform weak convergence of any (suitably constructed) ALE to a bounded IC on a representative subclass of the system of neighboring distributions Q_n—those distributions induced by simple perturbations $Q_n(\zeta, t)$, see p. 126 (ibid.).
Without additional assumptions, this weak convergence however does not carry over to convergence of the risk for an unbounded loss function in general, i.e. uniform integrability fails on any proper neighborhoods shrinking arbitrarily fast; which can be seen along the lines of Ruckdeschel [25, Prop. 2.1].

Modification of the shrinking neighborhood setup We instead propose the following modification of the neighborhoods for finite n: Only realizations of $U_1, \ldots, U_n$ are permitted, where $\sum U_i < n/2$. More precisely, accounting for non-symmetric ψ, we introduce

$$\check{b} := \inf \psi, \qquad \hat{b} = \sup \psi, \qquad \bar{b} := \tfrac{1}{2}(\hat{b} - \check{b}), \qquad \delta_0 := \tfrac{|\check{b}+\hat{b}|}{\min((-\check{b}),\hat{b})} \geq 0 \qquad (2.1)$$

and recall that in our situation, both the functional (Huber [14, (2.39),(2.40)]) and the finite sample (ε-contamination) breakdown point (Donoho and Huber [6, section 2.2]) of T respectively S_n are

$$\varepsilon_0 = 1/(2 + \delta_0) = \sup |\psi|/(\hat{b} - \check{b}) \qquad (2.2)$$

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

$$Q_n = \mathcal{L}\Big\{[(1 - U_i)X_i^{\rm id} + U_i X_i^{\rm di}]_i \,\Big|\, \sum U_i \leq \ulcorner \varepsilon_0 n \urcorner - 1\Big\} \qquad (2.3)$$

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 the *unmodified* MSE risk

$$R_n(S_n, r; \varepsilon_0) := \sup_{Q_n \in \tilde{Q}_n(r;\varepsilon_0)} n \int |S_n - \theta_0|^2 \, dQ_n \qquad (2.4)$$

Asymptotic negligibility of this modification The effect of this modification is negligible asymptotically: By the Hoeffding bound (B.1),

$$P(\textstyle\sum U_i \geq n\varepsilon_0) \leq \exp\Big(-2n(\varepsilon_0 - r/\sqrt{n}\,)^2\Big) \qquad (2.5)$$

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

Remark 2.1 (a) Thinning out the neighborhoods is equally relevant for the interchange of integration and maximization in the context of neighborhoods to a fixed radius ε: Replacing $r/\sqrt{n}$ by ε, asymptotic negligibility (2.5) continues to hold, as long as $\varepsilon < \varepsilon_0$, while the failure of uniform integrability persists.
(b) M-estimators have the well-known feature that in general the procedure with optimal efficiency [minimax MSE in our context] does not attain maximal breakdown point [works with minimally thinned out neighborhoods]; but just as already mentioned in Rousseeuw [23] and similarly as worked out in Yohai [32], both goals may be achieved simultaneously combining a starting M-estimator of maximal breakdown point with a correction by a one-/k-step construction with the f-o-o (or s-o-o, t-o-o) IC.

3 Main Theorem

Notation To $\psi:\mathbb{R}\to\mathbb{R}$ monotone let $\psi_t(x) := \psi(x-t)$ and $\psi_t^0 := \psi_t - \mathrm{E}\,\psi_t$ define the following functions

$$L(t) = \mathrm{E}\,\psi_t,\quad V(t)^2 = \mathrm{E}(\psi_t^0)^2,\quad \rho(t) = \mathrm{E}(\psi_t^0)^3 V(t)^{-3},\quad \kappa(t) = \mathrm{E}(\psi_t^0)^4 V(t)^{-4} - 3 \tag{3.1}$$

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

$$\psi(\check{y}_n) = \inf\psi + \mathrm{o}(\tfrac{1}{n^\gamma}),\qquad \psi(\hat{y}_n) = \sup\psi + \mathrm{o}(\tfrac{1}{n^\gamma}) \tag{3.2}$$

For $H \in \mathcal{M}_1(\mathbb{B}^n)$ and an ordered set of indices $I = (1 \le i_1 < \ldots < i_k \le n)$ denote H_I the marginal of H with respect to I.

Definition 3.1 *Consider sequences c_n, d_n, and κ_n in $\mathbb{R}$, in $(0,\infty)$, and in $\{1,\ldots,n\}$, respectively. We say that $(H^{(n)}) \subset \mathcal{M}_1(\mathbb{B}^n)$ 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 $i_n \le \kappa_n$*

$$1 - H_{I_n}^{(n)}((-\infty; c_n]^{i_n}) = \mathrm{o}(d_n)\qquad \Big[\,1 - H_{I_n}^{(n)}((c_n,\infty)^{i_n}) = \mathrm{o}(d_n)\,\Big] \tag{3.3}$$

General assumptions in this paper

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

(D) For some $\delta \in (0,1]$, L, V, ρ, and κ from (3.1) allow the expansions

$$L(t) = l_1 t + \tfrac{1}{2} l_2\, t^2 + \tfrac{1}{6} l_3\, t^3 + \mathrm{O}(t^{3+\delta}),\quad V(t) = v_0(1 + \tilde{v}_1\, t + \tfrac{1}{2}\tilde{v}_2\, t^2) + \mathrm{O}(t^{2+\delta}) \tag{3.4}$$

$$\rho(t) = \rho_0 + \rho_1\, t + \mathrm{O}(t^{1+\delta}),\qquad \kappa(t) = \kappa_0 + \mathrm{O}(t^\delta) \tag{3.5}$$

(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

$$\lim_{t_0\to 0}\limsup_{s\to\infty}\sup_{|t|\le t_0}|f_t(s)| < 1 \tag{3.6}$$

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 3.5 is possible, however, and given in Ruckdeschel [25].

(c) For an expansion of the MSE up to $\mathrm{o}(n^{-1/2})$, the κ part of assumption (3.5) can be dropped, and we may use assumptions

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

$$L(t) = l_1 t + l_2/2\, t^2 + \mathrm{O}(t^{2+\delta}),\qquad V(t) = v_0(1 + \tilde{v}_1\, t) + \mathrm{O}(t^{1+\delta}),\qquad \rho(t) = \rho_0 + \mathrm{O}(t^\delta) \tag{3.7}$$

(C') There exist $t_0 > 0$, $s_0 > 0$ such that for all $s_1 > s_0$

$$\hat{f}_{s_0,t_0}(s_1) := \sup_{s_0\le s\le s_1}\sup_{|t|\le t_0}|f_t(s)| < 1 \tag{3.8}$$

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.

Illustration We specialize the assumptions for $F = \mathcal{N}(0,1)$, i.e. $\Lambda_f(x) = x$, and $\psi(x) = \hat\eta_c(x) = A_c x \min\{1, c/|x|\}$ from (1.6) with A_c such that $\hat\eta_c \in \Psi$:

Proposition 3.4 *For $F = \mathcal{N}(0,1)$ and for $\psi = \eta_c$ an influence curve to $c \in (0,\infty)$ of Hampel-form $\eta_c = A_c(x \min\{1, c/|x|\}$ with $A_c = (2\Phi(c)-1)^{-1}$, assumptions (bmi) to (C) are in force; in particular the bounds in (Lb) and (Vb) are even exponential. With $\Phi(x)$ the c.d.f. of $\mathcal{N}(0,1)$ and $\varphi(x)$ its density, we obtain $l_2 = 0$, $\tilde v_1 = 0$, $\rho_0 = 0$. For $c \in (0,\infty)$, we get*

$$l_3 = \frac{2c\varphi(c)}{(2\Phi(c)-1)}, \quad v_0^2 = 2b^2(1-\Phi(c)) + A_c(1-2b\varphi(c)), \quad \tilde v_2 = \frac{6\Phi(c) - 4\Phi(c)^2 - 2 - 2c\varphi(c)}{2c^2(1-\Phi(c)) + 2\Phi(c) - 1 - 2c\varphi(c)}$$
$$\rho_1 = \frac{3A_c^3\,(1 - 2\Phi(c) + 2c\varphi(c))}{v_0^3} + 3v_0^{-1}, \quad \kappa_0 = \frac{2c^4\,(1-\Phi(c)) - 2c(c^2+3)\varphi(c) + 3(2\Phi(c)-1)}{[2c^2\,(1-\Phi(c)) + 2\Phi(c) - 1 - 2c\varphi(c)]^2} - 3$$

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

Theorem 3.5 (Main Theorem) *In our one-dim. location model assume (bmi) to (C)*
(a) *the maximal MSE of the M-estimator S_n to scores-function ψ expands to*

$$R_n(S_n, r, \varepsilon_0) = r^2 b^2 + {v_0}^2 + \tfrac{r}{\sqrt{n}}\,A_1 + \tfrac{1}{n}\,A_2 + \mathrm{o}(n^{-1}) \tag{3.9}$$

with

$$A_1 = {v_0}^2\Big(\pm(4\,\tilde v_1 + 3\,l_2\,)b + 1\Big) + b^2 + [2\,b^2 \pm l_2\,b^3\,]\,r^2 \tag{3.10}$$
$$\begin{aligned} A_2 = {v_0}^3\Big((l_2 + 2\,\tilde v_1\,)\rho_0 + \tfrac{2}{3}\,\rho_1\Big) + {v_0}^4\,(3\,\tilde v_2 + \tfrac{15}{4}\,l_2^2 + l_3 + 9\,\tilde v_1^2 + 12\,\tilde v_1\,l_2\,) + \\ +[\,{v_0}^2\Big((3\,\tilde v_2 + 3\,\tilde v_1^2 + \tfrac{15}{2}\,l_2^2 + 2\,l_3 + 12\,\tilde v_1\,l_2\,)b^2 + 1 \pm (8\,\tilde v_1 + 6\,l_2\,)\,b\Big) + \\ \pm 3\,l_2\,b^3 + 5\,b^2\,]\,r^2 + \Big((\tfrac{5}{4}\,l_2^2 + \tfrac{1}{3}\,l_3\,)b^4 \pm 3\,l_2\,b^3 + 3\,b^2\Big)\,r^4 \end{aligned} \tag{3.11}$$

and we are in the $-$ [+]*-case depending on whether* (3.12) *or* (3.13) *below applies.*
(b) *let $P_n^{\rm di} := \bigotimes_{i=1}^n P_{n,i}^{\rm di}$ be contaminating measures for* (1.2). *Then Q_n with $P_n^{\rm di}$ as contaminating measures generates maximal risk in* (3.9) *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*

$$(P_n^{\rm di}) \text{ is } K_1(n)\text{–concentrated left of } \check y_n - b\sqrt{k_2\log(n)/n} \text{ up to } \mathrm{o}(n^{-1}) \tag{3.12}$$

or

$$(P_n^{\rm di}) \text{ is } K_1(n)\text{–concentrated right of } \hat y_n + b\sqrt{k_2\log(n)/n} \text{ up to } \mathrm{o}(n^{-1}) \tag{3.13}$$

More precisely, if $\sup\psi$ < [>] $-\inf\psi$, the maximal MSE is achieved by contaminations according to (3.12) *[*(3.13)*]. In case $\sup\psi = -\inf\psi$,* (3.12) *[*(3.13)*] applies if*

$$\tilde v_1 \;>\; [<] -\tfrac{l_2}{4}\Big(\tfrac{b^2}{v_0^2}(r^2+3)(1 + \tfrac{r}{\sqrt{n}} - \tfrac{2r^2}{n}) + 3(1 - \tfrac{b^2}{v_0^2})\Big) \tag{3.14}$$

If $\sup\psi = -\inf\psi$ and there is "=" in (3.14), (3.12) *and* (3.13) *generate the same risk up to order* $\mathrm{o}(n^{-1})$.

Remark 3.6 (a) A sufficient condition for (3.12)/(3.13) is that $P_n^{\rm di}$ is concentrated strictly right [left] of $\hat y_n + b\sqrt{2\log(n)/n}$ $[\check y_n - b\sqrt{2\log(n)/n}\,]$.

(b) An almost necessary condition for (3.12)/(3.13) to achieve maximal risk is that $P_n^{\rm di}$ is concentrated strictly right [left] of $\hat y_n - b\sqrt{2\log(n)/n}$ $[\check y_n + b\sqrt{2\log(n)/n}\,]$.

(c) 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.

(d) As announced, for $r = 0$, the approximation is one order faster than under contamination.

(e) **The maximal MSE on $\tilde Q_n$ is always underestimated by f-o asymptotics,** as maximality always forces A_1 to be non-negative.

(f) Let Q_n^0 be any distribution in $\tilde Q_n$ attaining maximal risk in Theorem 3.5. Under symmetry or more specifically if $l_2 = v_1 = \rho_0 = 0$, (3.9) becomes

$$n\,{\rm E}_{Q_n^0}[S_n^2] = \big(r^2b^2 + v_0{}^2\big)\Big(1 + \tfrac{r}{\sqrt n}\Big) + \tfrac{r}{\sqrt n}\,\big(b^2(1+r^2)\big) + {\rm O}(n^{-1}) \tag{3.15}$$

(g) **Relevance for the fixed neighborhood approach:** If you consider the fixed neighborhood approach (of radius ε) and formally plug in $r = \varepsilon\sqrt n$ into (3.9), you obtain the following approximation for the unstandardized maximal MSE on the thinned out (fixed-radius) neighborhood:

$$\begin{aligned}{\rm MSE}(S_n,\varepsilon,\varepsilon_0) &= \varepsilon^2 b^2 + \varepsilon^3\,[2\,b^2 \pm l_2\,b^3\,] + \varepsilon^4\Big(\big(\tfrac54\,l_2^2 + \tfrac13\,l_3\,\big)b^4 \pm 3\,l_2\,b^3 + 3\,b^2\Big) +\\ &+\frac1n v_0^2 + \frac{\varepsilon}{n}\Big[v_0{}^2\,\Big(\pm(4\,\tilde v_1 + 3\,l_2\,)b + 1\Big) + b^2\Big] + \frac{\varepsilon^2}{n}\Big[5\,b^2 \pm 3\,l_2\,b^3 +\\ &+v_0{}^2\Big((3\tilde v_2 + 3\tilde v_1^2 + \tfrac{15}{2}l_2^2 + 2l_3 + 12\tilde v_1 l_2)b^2 + 1 \pm (8\tilde v_1 + 6l_2)\,b\Big)\Big] + R_n\end{aligned} \tag{3.16}$$

for some remainder R_n the order of which however is uncertain; it should be valid for small ε, and is at least of order ${\rm O}(1/n^2) + {\rm O}(\varepsilon^5)$. These terms once more show that for the fixed-neighborhood approach, already for moderate sample sizes, bias becomes dominant, i.e.; in our case, we end up with the median as optimal procedure.

3.1 Cross-checks

3.1.1 Check with results by Fraiman et al.

In the symmetric case, the first cross check comes with the asymptotic formula for variance ${\rm asVar}(\psi)$ and (maximal) bias $B(\psi) := {\rm asBias}(\psi)$ as to be found in Fraiman et al. [8], where we have to identify $\varepsilon = r/\sqrt n$. Here, ${\rm asBias}(\psi)/\sqrt n$ is defined as zero β of $\beta \mapsto (1-\varepsilon)\int \psi_\beta\,dF + \varepsilon b$, and ${\rm asVar}(\psi) := V_1/V_2^2$ for $V_1 = (1-\varepsilon)\int \psi^2_{B(\psi)}\,dF + \varepsilon b^2$ and $V_2 = (1-\varepsilon)\int \dot\psi_{B(\psi)}\,dF$. Assuming that $\int \dot\psi_{B(\psi)}\,dF = L'(B(\psi))$ and using that $\int \psi_{B(\psi)}\,dF = -B(\psi) + {\rm o}(B^2))$, $\int \psi^2_{B(\psi)}\,dF = V(B(\psi))^2 + L(B(\psi))^2 = v_0^2(1+{\rm o}(B))$, $L'(B(\psi))^2 = -1 + {\rm o}(B)$, we get that

$${\rm asBias}(\psi) = \sqrt n\,b\varepsilon(1 + \varepsilon + {\rm o}(\varepsilon)) = rb(1 + \tfrac{r}{\sqrt n} + {\rm o}(n^{-1/2})) \tag{3.17}$$

$${\rm asVar}(\psi) = (1+\varepsilon)v_0^2 + \varepsilon b + {\rm o}(\varepsilon) = v_0^2 + \tfrac{r}{\sqrt n}(v_0^2 + b) + {\rm o}(n^{-1/2}) \tag{3.18}$$

and hence —in accordance with formula (3.9)—

$${\rm asMSE}(\psi) = (v_0^2 + r^2b^2)(1 + \tfrac{r}{\sqrt n}) + \tfrac{r}{\sqrt n}b^2(1+r^2) + {\rm o}(n^{-1/2}) \tag{3.19}$$

3.1.2 Check with higher order asymptotics for the median

The second check comes with the higher asymptotics for the median from Ruckdeschel [25]. In a first step, we assume that with $f_0 > 0$ and some $\delta \in (0,1]$,

$$f(t) = f_0 + f_1 t + \mathrm{O}(t^{1+\delta}) \qquad (3.20)$$

As for the median, $\psi_{\mathrm{Med}} = \mathrm{sign}(x)/(2 f_0)$, we have $v_0 = b = \frac{1}{2f_0}$ and $\varepsilon_0 = 1/2$. For the moment we ignore the fact, that conditions (C)/(C') are not fulfilled. Easy calculations give $l_2 = -f_1/f_0$, $\tilde v_1 = 0$, $\rho_0 = 0$, so that with our formula (3.9) we obtain for odd sample size n

$$R_n(\psi_{\mathrm{Med}_n}, r, \tfrac{1}{2}) = \frac{1}{4f_0^2}\Big((1+r^2)[1+\tfrac{2r}{\sqrt{n}}] - \tfrac{r}{\sqrt{n}}\tfrac{f_1}{2f_0^2}(r^2+3)\Big) + \mathrm{o}(n^{-1/2}) \qquad (3.21)$$

in complete agreement with Ruckdeschel [25]. As a next step we compare this to t-o asymptotics to be obtained by (3.9)—again ignoring condition (C). We get $l_3 = -f_2/f_0$, $\tilde v_2 = -4f_0^2$, $\rho_1 = 4f_0$, and hence for odd sample size n, after some reordering

$$\begin{aligned} R_n(\psi_{\mathrm{Med}_n}, r, \tfrac{1}{2}) \stackrel{?}{=} \mathrm{o}(\tfrac{1}{n}) + \tfrac{1}{4f_0^2}\Big\{&(1+r^2) + \tfrac{r}{\sqrt{n}}\Big(2(1+r^2) + \tfrac{r^2+3}{2}\tfrac{|f_1|}{f_0^2}\Big) + \\ &+ \tfrac{1}{n}\Big(\boxed{\tfrac{4}{3}} - 3 + 3r^2 + 3r^4 + \tfrac{3r^2(3+r^2)}{2}\tfrac{|f_1|}{f_0^2} - \tfrac{3+6r^2+r^4}{12}\tfrac{f_2}{f_0^3} + \tfrac{5(3+6r^2+r^4)}{16}\tfrac{f_1^2}{f_0^4}\Big)\Big\} \end{aligned} \qquad (3.22)$$

and it is just the framed term $\frac{4}{3}$, which is coming in as $\frac{2}{3}\rho_1 v_0$ from (3.11), which causes a difference to the result of Ruckdeschel [25], where we get the value 1 instead. This discrepancy, however, is in fact due to the failure of condition (C), because Theorem B.2, which we need to prove (3.9), is not available in this case.

4 Relations to other approaches

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, cf. among others Ibragimov and Linnik [16], Bhattacharya and Rao [3], Pfanzagl [20], Hall [10], Barndorff-Nielsen and Cox [2] and Taniguchi and Kakizawa [30].

Asymptotic expansions of the moments of statistical estimators —like MSE in our case— have already been studied by Gusev [9] and Pfaff [19]; 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, cf. Daniels [5], Hampel [11], Field and Ronchetti [7].

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 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 (2.3). 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 apply Field and Ronchetti [7, Theorem 4.3] which at least covers the Hampel-type solutions. The pointwise formulation of assumption A4.2 therein, i.e.; there exists an open subset $U \subset \mathbb{R}$, such that (i) for each $\theta \in \mathbb{R}$, $F(U - \theta) = 1$ and (ii) $D\psi$, $D^2\psi$, $D^3\psi$ exist on U, seems problematic, however, 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 (3.9), 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 A.3, we extract an argument to be expanded from the exponential, which also is an idea behind the saddlepoint approximation, cf. Field and Ronchetti [7, p. 26]. On the other hand, even in the restricted neighborhoods of (2.3), 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 [26].

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.

5.1 Simulation design

Under `R 2.11.0`, cf. R Development Core Team [21], 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

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

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 (1.6) with clipping heights $c = 0.5, 0.7, 1, 1.5, 2$ and $c_0(r)$, the f-o-o clipping height according to (1.6). All empirical MSE's come with asymptotic 95%–confidence intervals, which are based on the CLT for the variables

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

Note that with respect to (3.12)/(3.13), and the considered estimators, a contamination point 100 will largely suffice to attain the maximal MSE on $\tilde{Q}_n$.

5.2 Numerical evaluations

By means of relations (1.9) 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 Ruckdeschel and Kohl [27]. This algorithm is based on FFT resp. discrete Fourier Transformation (DFT) and is implemented in `R` within the package `distr` available on `CRAN`, see Ruckdeschel et al. [28], Ruckdeschel et al. [29].

In Ruckdeschel and Kohl [26], 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 $\tilde{Q}_n$.

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 [26].

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 (1.3) no longer are independent. So we may only apply the FFT-based Algorithm from Ruckdeschel and Kohl [26] to an absolutely continuous inner part and have to calculate the rest by explicitly summing up the events—for details see the cited reference and the `R`-program available on the web-page to this article.

On the other side, Algorithm D uses the fact that by the exponential negligibility shown in subsection 2, 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 Ruckdeschel and Kohl [27].

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.

Table 1 emp., num., and as. MSE at $r = 0.1$, $c = 0.7$

n/ situation		simulation $\bar{S}_n$	[low;	up]	numeric Algo C	Algo D	asymptotics n^0	$n^{-1/2}$	n^{-1}
5	id	1.147	[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

Table 2 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

5.3.1 Fixed procedure, fixed radius

To get an idea of the speed of the convergence of the MSE to its asymptotic values, we consider the `H07`-estimator from Andrews et al. [1], 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 3.5. 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 1/2. In Table 3 we consider the relative MSE, calculated as the quotient $\mathrm{MSE}(c, r)/\mathrm{MSE}(c_0(r), r)$. 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 asymptotic terms derived by means of Theorem 3.5. We already recognize a very good approximation down to very small sample sizes.

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

Table 3 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 4 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

tabulated in Table 4. 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 5 and 6. The simulations already indicate that the answer should be affirmative. The numeric and asymptotic values for the median are taken from Ruckdeschel [25]. 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

Table 5 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 6 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

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 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 7, 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 1E − 5 which may result in a deviation from the "real" n_0 of ±2 for $n_0 < 200$.

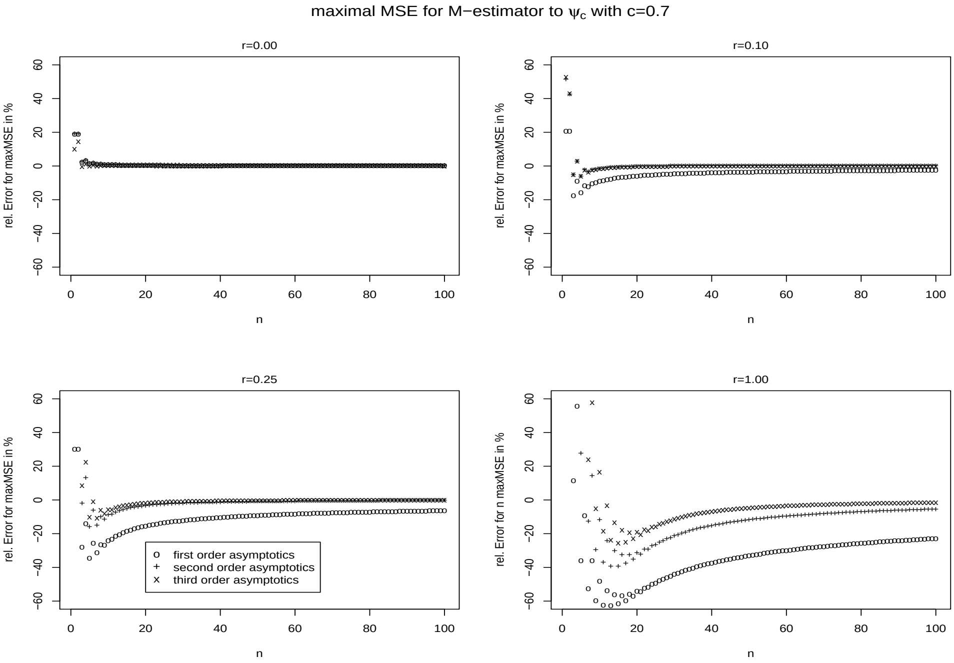

Fig. 1 The mapping $n \mapsto \text{rel.error}(\text{MSE}_n(\psi_c))$ for $c = 0.7$ and $F = \mathcal{N}(0, 1)$.

Table 7 Minimal n_0 such that for $n \geq n_0$ the relative error using first to third order asymptotics for approximating $\text{MSE}_n(\psi_c)$ 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 $\text{O}(n^{-3/2})$, 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 $-3\text{E}-3$, $-7\text{E}-4$, $-3\text{E}-4$, $-2\text{E}-2$, $-2\text{E}-2$, $-1.3\text{E}-1$, and $-2\text{E}-4$.

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 $\tilde{Q}_n(r, \varepsilon_0)$ from (2.3) to

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

for some fixed ε_0' such that

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

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

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

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

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

Proposition 6.1 *In the location model of Subsection 1.1, assume (bmi), (D), and (C) from section 3; additionally assume that the central distribution F satisfies* (6.4). *Then, on $\tilde{Q}_n'(r;\varepsilon_0')$, the assertions of Theorem 3.5 —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 of Subsection 1.1, assume: For any $d > 0$,*

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

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 (3.12) resp. (3.13) almost characterize the risk-maximizing contaminations:

Proposition 6.3 *Under the assumptions of Theorem 3.5, let $\delta_0, c_0 > 0$. Assume that $\hat{b} = b$ and let $B_n := \inf\{x \,\big|\, \psi(x) \ge b - c_0/\sqrt{n}\}$. Assume that, for $K = \sum_{i=1}^n U_i$ and $k > (1-\delta)r\sqrt{n}$,*

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

Then, eventually in n, for any such sequence of contaminations $Q_n^\flat \in \tilde{Q}(r)$, the maximal MSE as in condition (3.13) *(i.e. with positive bias) in* (3.9) *cannot be attained. More precisely,*

$$R_n(S_n, r) - n\,\mathrm{E}_{Q_n^\flat} S_n^2 \ge 2p_0v_0(rc_0 + b)/(n\sqrt{2\pi}) \tag{6.7}$$

A corresponding relation holds for condition (3.12).

6.2 Convergence of variance and bias separately

The technique used to derive Theorem 3.5 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 $Q_n^0 \in \tilde{Q}_n(r;\varepsilon_0)$ according to* (3.12) *resp.* (3.13) *admits the following expansions*

$$\sqrt{n}\left|\mathrm{Bias}(S_n, Q_n^0)\right| = \left| rb + \tfrac{1}{\sqrt{n}}\, B_{1,0} + \tfrac{r^2}{\sqrt{n}}\, B_{1,1} + \tfrac{r}{n}\, B_2 \right| + \mathrm{o}(n^{-1}) \tag{6.8}$$

$$n\,\mathrm{Bias}^2(S_n, Q_n^0) = r^2 b^2 + \tfrac{r}{\sqrt{n}}\, C_1 + \tfrac{1}{n}\, C_2 + \mathrm{o}(n^{-1}) \tag{6.9}$$

$$n\,\mathrm{Var}(S_n, Q_n^0) = v_0^2 + \tfrac{r}{\sqrt{n}}\, D_1 + \tfrac{1}{n}\, D_2 + \mathrm{o}(n^{-1}) \tag{6.10}$$

with

$$B_{1,0} = (\tfrac{1}{2} l_2 + \tilde{v}_1) v_0^2, \qquad B_{1,1} = b(1 \pm \tfrac{1}{2} l_2 b) \tag{6.11}$$

$$\begin{aligned} B_2 = {} & \Big[(\tfrac{1}{2} l_2^2 + \tfrac{1}{6} l_3) b^3 + b \pm l_2 b^2\Big] r^2 + b(1 \pm \tfrac{1}{2} l_2 b) + \\ & + \Big[(\tfrac{1}{2} l_3 + \tfrac{3}{2} l_2^2 + \tilde{v}_2 + \tilde{v}_1^2 + 3\,\tilde{v}_1\, l_2) b \pm \tfrac{1}{2} l_2 \pm \tilde{v}_1\Big] {v_0}^2 \end{aligned} \tag{6.12}$$

$$C_1 = b^2 r^2 (\pm l_2 b + 2) \pm b(l_2 + 2\tilde{v}_1) v_0^2 \tag{6.13}$$

$$\begin{aligned} C_2 = {} & (\tilde{v}_1\, l_2 + \tfrac{1}{4}\, l_2^2 + \tilde{v}_1^2) {v_0}^4 + \Big[3\, b^2 \pm 3\, l_2\, b^3 + (\tfrac{5}{4}\, l_2^2 + \tfrac{1}{3}\, l_3) b^4\Big] r^4 + \\ & + \Big[(\tfrac{7}{2}\, l_2^2 + l_3 + 2\,\tilde{v}_2 + 2\,\tilde{v}_1^2 + 7\,\tilde{v}_1\, l_2) b^2\, {v_0}^2 \pm (2\, l_2 + 4\,\tilde{v}_1)\, b\, v_0^2 + 2b^2 \pm l_2\, b^3\Big] r^2 \end{aligned} \tag{6.14}$$

$$D_1 = \Big[\pm 2(\, l_2 + \tilde{v}_1) b + 1\Big] {v_0}^2 + b^2 \tag{6.15}$$

$$\begin{aligned} D_2 = {} & (l_3 + \tfrac{7}{2}\, l_2^2 + 11\,\tilde{v}_1\, l_2 + 8\,\tilde{v}_1^2 + 3\,\tilde{v}_2) {v_0}^4 + \Big(\tfrac{2}{3}\,\rho_1 + (l_2 + 2\,\tilde{v}_1)\rho_0\Big) {v_0}^3 + \\ & \Big[\Big((l_3 + \tilde{v}_1^2 + \tilde{v}_2 + 5\,\tilde{v}_1\, l_2 + 4\, l_2^2) b^2 \pm 4(l_2 + \tilde{v}_1) b + 1\Big) {v_0}^2 \pm 2\, l_2\, b^3 + 3b^2\Big] r^2 \end{aligned} \tag{6.16}$$

where we are in the $-$ [+]*-case according to whether* (3.12) *or* (3.13) *applies.*

For a proof to this proposition, we may proceed exactly as in the proof of Theorem 3.5; only in (A.38), 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`.

A Proofs

A.1 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 $d_i = \psi^{-1}(c_i)$

$$P(\psi_t(X^{\mathrm{id}}) \in (c_1, c_2)) = P(d_1 + t < X^{\mathrm{id}} < t + d_2) = \int_{d_1+t}^{d_2+t} dF \tag{A.1}$$

But

$$\int_{d_1+t}^{d_2+t} dF = G_0(c_2) - G_0(c_1) + \mathrm{o}(t^0) \tag{A.2}$$

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 (3.6) holds. □

A.2 Proof to Proposition 3.4

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

$$L_c(t) = A[c-(c+t)\Phi(t+c)+(t-c)\Phi(t-c)+\varphi(t-c)-\varphi(t+c)], \quad L_\infty(t) = -t, \quad L_0(t) = \sqrt{\tfrac{\pi}{2}}\,(1-2\Phi(t)) \tag{A.3}$$

all arbitrarily often differentiable functions, so the l_i-part of (D) holds 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, $W = S - L^2$, $\quad W' = S' - 2LL'$, $\quad W'' = S'' - 2L'^2 - 2LL''$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 $V(t) = \sqrt{W_0}(1 + \frac{\tilde{W}_1 t}{2} + \frac{(2\tilde{W}_2 - \tilde{W}_1^2)t^2}{8}) + \mathrm{O}(t^{2+\delta})$ 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 $0 < c < \infty$

$$S(t) = A_c^2\Big[c^2(1 - \Phi(t+c) + \Phi(t-c)) + (1+t^2)(\Phi(t+c) - \Phi(t-c)) + (t-c)\varphi(t+c) - (t+c)\varphi(t-c)\Big]$$

and $S(t) = 1 + t^2 \quad$ for $c = \infty$, $\qquad S(t) = \frac{\pi}{2} = b^2 \quad$ for $c = 0$, so (3.4) 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 $\delta > 0$

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

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

$$V^2(t) = b^2 - (b + \mathrm{o}(\exp(-t^2/2)))^2 = \mathrm{o}(\exp(-t^2/2))$$

For $\rho(t)$ and $\kappa(t)$, we introduce $M(t) := \mathrm{E}[\psi(X-t)^3]$, $N(t) := \mathrm{E}[\psi(X-t)^4]$. 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 $\rho_0 = v_0^{-3}M(0)$, $\kappa_0 = V^{-4}N(0) - 3$. 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 $c = \infty$, $M(t) = -3t - t^3$, $M'(t) = -3 - 3t^2$, $N(t) = t^4 + 6t^2 + 3$and for $c = 0$, $M(t) = (\sqrt{\frac{\pi}{2}})^3(1 - 2\Phi(t))$, $M'(t) = -2(\sqrt{\frac{\pi}{2}})^3\varphi(t)$, $N(t) = \frac{\pi^2}{4}$, while for $0 < c < \infty$

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

This gives the assertion as to ρ_0, ρ_1 and κ_0, and hence (3.5) holds. 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 Lemma 3.2 entails (C). □

A.3 Proof of Theorem 3.5

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 (1.8).

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

$$D_{k,\tilde{t}} := \{K = k, \tilde{T}_{n,k}(\sqrt{t}\,) = \tilde{t}\,\} \tag{A.4}$$

Thus

$$n\,\mathrm{MSE}(S_n, Q_n \,|\, D_{k,\tilde{t}}\,) = \int_0^\infty \Pr(S_n^2 \geq t \,|\, D_{k,\tilde{t}}\,)\,dt = \int_0^\infty \Pr(S_n \geq \sqrt{t} \,|\, D_{k,\tilde{t}}\,)\,dt + \int_0^\infty \Pr(S_n \leq -\sqrt{t} \,|\, D_{k,\tilde{t}}\,)\,dt \tag{A.5}$$

For the sequel, we define $\bar{n} := n - k$, $s_{n,k} := s_{n,k}(t) = \frac{-\tilde{t} - \bar{n}L(t)}{\sqrt{\bar{n}}\;V(t)}$. 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):

	$K < k_1 r\sqrt{n}$	$k_1 r\sqrt{n} \leq K < \varepsilon_0 n$	$K \geq \varepsilon_0 n$
$\lvert t\rvert \leq k_2 b^2 \log(n)/n$	(I)	(II)	excluded
$k_2 b^2 \log(n)/n < \lvert t\rvert \leq Cn^{1+3/\delta}$	(III)	(II)	excluded
$\lvert t\rvert > Cn^{1+3/\delta}$	(IV)	(IV)	excluded

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 (3.9) stated in Theorem 3.5.

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

$$\Pr\{S_n \leq t \,\Big|\, K = k\,\} + R_n^{(0)}(k) = \Pr(T_{n,k}(t) < -\tilde{T}_{n,k}(t)) = \Pr\Big(\frac{T_{n,k}(t) - \bar{n}L(t)}{\sqrt{\bar{n}}V(t)} < -\frac{\tilde{T}_{n,k}(t) - \bar{n}L(t)}{\sqrt{\bar{n}}V(t)}\Big) \tag{A.6}$$

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

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

$$\Pr\{S_n \le t\,|D_{k,\tilde{t}}\} + \tilde{R}_n^{(0)}(k,\tilde{t}) = \Pr\Big(\frac{T_{n,k}(t) - \bar{n}L(t)}{\sqrt{\bar{n}}\,V(t)} < s_{n,k}(t)\Big) \tag{A.7}$$

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

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 $t > t_1^2$

$$\Pr\{S_n > \sqrt{t}\,\Big|\,D_{k,\tilde{t}}\} \le \Pr\Big(T_{n,k}(\sqrt{t}\,) - \bar{n}L(\sqrt{t}\,) \ge -\tilde{t} - \bar{n}L(\sqrt{t}\,)\Big) \overset{\text{Cheb.}}{\le} \frac{\bar{n}V^2(\sqrt{t}\,)}{(\tilde{t} + \bar{n}L(\sqrt{t}\,))^2} \overset{\text{(Vb)}}{\le} \frac{\bar{n}C' t^{-(1+\delta)}}{(\tilde{t} + \bar{n}L(\sqrt{t}\,))^2} \le$$

$$\le \frac{nC' t^{-(1+\delta)}}{(\tilde{t} + \bar{n}\check{b} + \eta)^2} \overset{\tilde{t} \le k\hat{b}}{\le} \frac{nC' t^{-(1+\delta)}}{[k\hat{b} + \bar{n}\check{b} + \eta]^2} = \frac{nC' t^{-(1+\delta)}}{[k(\hat{b} - \check{b}) + n\check{b} + \eta]^2} \overset{k \le \varepsilon_0 n}{\le} \frac{nC' t^{-(1+\delta)}}{(\check{b} - \eta)^2} \tag{A.8}$$

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

$$\frac{C' n^2}{(b - \eta)^2} \int_{Cn^{1+3/\delta}}^{\infty} t^{-(1+\delta)}\,dt = \frac{C' C^{-\delta} n^{-1-\delta}}{\delta(\check{b} - \eta)^2} = \mathrm{o}(n^{-1}) \tag{A.9}$$

Negligibility of case (II)

Lemma A.1 *Let*

$$\kappa := k_1 \log k_1 + 1 - k_1 \tag{A.10}$$

Then it holds that

$$\Pr(\mathrm{Bin}(n, r/\sqrt{n}\,) > k_1 r \sqrt{n}) \le \exp\big(-\kappa\, r \sqrt{n} + \mathrm{o}(\sqrt{n}\,)\big) \tag{A.11}$$

Proof Ruckdeschel [25, Lem. A.2] ◻

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 A.1 the contribution of (II) is indeed $\mathrm{o}(n^{-1})$.

Another consequence of the exponential decay of (A.11) is that we may neglect values of $K > k_1(n) r \sqrt{n}$ when integrating along K.

Corollary A.2 *Let* $K \sim \mathrm{Bin}(n, r/\sqrt{n}\,)$*. Then, in the setup of Lemma A.1, for any* $j \in \mathbb{N}$*,*

$$\mathrm{E}[K^j\,\mathrm{I}_{\{X \ge k_1(n) r \sqrt{n}\}}] = \mathrm{o}(e^{-rn^d}) \tag{A.12}$$

for any $0 < d < \sqrt{n}$*.*

Proof $\mathrm{E}[K^j\,\mathrm{I}_{\{K \ge k_1(n) r \sqrt{n}\}}] \le n^j \Pr(X > k_1(n) r \sqrt{n}) \overset{\text{(A.11)}}{=} \mathrm{o}(e^{-rn^d})$. ◻

Negligibility of case (III) We apply Hoeffding's first bound from Lemma B.1:

$$\Pr\{S_n > \sqrt{t}\,\Big|\,D_{k,\tilde{t}}\} \le \Pr(T_{n,k}(\sqrt{t}\,) \ge -\tilde{t}\,\Big|\,D_{k,\tilde{t}}\,) \le \exp(-2n\Delta^2/b^2) \tag{A.13}$$

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

$$L(\sqrt{t}\,) \le L(b\sqrt{k_2 \log(n)/n}\,) = -b\sqrt{k_2 \log(n)/n} + \mathrm{o}(\sqrt{\log(n)/n}\,) \tag{A.14}$$

Thus

$$\Delta \ge -L(\sqrt{t}\,) - \frac{kb}{n} \overset{\text{(A.14)}}{>} \frac{b}{\sqrt{n}}[\sqrt{k_2 \log(n)} + \mathrm{o}(\sqrt{\log(n)}\,)] \tag{A.15}$$

and $\exp(-2\frac{n\Delta^2}{b^2}) < n^{-2k_2}(1 + \mathrm{o}(n^0))$. This latter is $\mathrm{o}(n^{-3-3/\delta})$ and thus integrating $n\,\mathrm{MSE}$ out along (III) we get something of order $\mathrm{o}(n^{-1})$.

Asymptotic normality On (I), by Lemma 1.1

$$\Pr\{S_n \geq \sqrt{t} \,\Big|\, D_{k,\bar{t}}\} = \Pr\left(\frac{T_{n,k}(\sqrt{t}) - \bar{n}L(\sqrt{t})}{\sqrt{\bar{n}}\, V(\sqrt{t})} > s_{n,k}(t) \right) + O(e^{-\gamma n}) \tag{A.16}$$

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 B.2(b) to (A.5)/(A.7), identifying

$$\xi_{i,t} := \frac{1}{V(t)}[\psi_t(X_{j_i}) - L(t)], \qquad i = 1, \ldots, \bar{n} \tag{A.17}$$

and setting $\Theta := \Theta_n = \{|t| \leq k_2 b^2 \log(n)/n\}$. 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 (B.7) of the theorem holds if n is large enough.
We note that if in Theorem 3.5, we limit ourselves to term A_1 and hence only assume (C'), we may apply Theorem B.2(a).
With $G_{n,t}(s)$ from (B.4) we define $\tilde{G}_{n,t}(u) := G_{n,t}(s_{n,k}(u))$, $\tilde{G}_n(t) := \tilde{G}_{n,t}(t)$ and obtain for $|t| \leq k_2 b^2 \log(n)/n$ and $K < k_1 r \sqrt{n}$ uniformly in t and k:

$$O(\exp(-\gamma n)) + \Pr\{S_n \geq \sqrt{t} \,\Big|\, D_{k,\bar{t}}\} = \Pr\big(\sum_{i=1}^{\bar{n}} \xi_{i,\sqrt{t}} > s_{n,k}(\sqrt{t})\big) = 1 - \tilde{G}_n(\sqrt{t}) + O(n^{-3/2}) \tag{A.18}$$

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

$$n^\natural = \sqrt{\bar{n}/n}, \qquad l_n = n^\natural \sqrt{k_2 \log(n)}, \qquad l_n^{(0)} = k_2 b^2 \log(n)/n \tag{A.19}$$

we obtain

$$\begin{aligned} n\, \mathrm{MSE}(S_n, Q_n \,\big|\, D_{k,\bar{t}}) &= (n^\natural)^{-2}\, \bar{n} \int_0^{l_n^{(0)}} 1 - \tilde{G}_n(\sqrt{t}) + \tilde{G}_n(-\sqrt{t})\, dt + o(n^{-1}) = \\ &= 2(n^\natural)^{-2} \int_0^{bl_n} u\Big(1 - \tilde{G}_n(\frac{u}{\sqrt{\bar{n}}}) + \tilde{G}_n(-\frac{u}{\sqrt{\bar{n}}})\Big)\, du + o(n^{-1}) \end{aligned} \tag{A.20}$$

As $\tilde{G}_n$ is arbitrarily smooth, integration by parts is available and gives

$$n\, \mathrm{MSE}(S_n, Q_n \,\big|\, D_{k,\bar{t}}) = R_n + (n^\natural)^{-2} \int_{-bl_n}^{bl_n} \frac{u^2}{\sqrt{\bar{n}}}\, G_n'(\frac{u}{\sqrt{\bar{n}}})\, du + o(n^{-1}) \tag{A.21}$$

with

$$R_n := k_2\, \log(n)\, b^2 \,[1 - \tilde{G}_n(b\, \sqrt{\tfrac{k_2 \log(n)}{n}}) - \tilde{G}_n(-b\, \sqrt{\tfrac{k_2 \log(n)}{n}})\,] \tag{A.22}$$

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

$$s_{n,k}(\pm b\, \sqrt{\tfrac{k_2 \log(n)}{n}}) \overset{(3.4)}{=} \frac{O(\sqrt{n}) \pm b\, \sqrt{\frac{k_2 \bar{n}^2 \log(n)}{n}} + O(\frac{\bar{n} \log(n)}{n})}{\sqrt{\bar{n}}\,(v_0 + o(n^0))} = \frac{\pm b\sqrt{k_2 \log(n)}}{v_0}(1 + o(n^0)) \tag{A.23}$$

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

$$|\tilde{s}_{n,k}(\pm b\, \sqrt{k_2 \log(n)})| > \sqrt{2 \log(n)} \tag{A.24}$$

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

$$\max\big(1 - G_{n,t}(s),\, G_{n,t}(-s)\big) \leq K|s|^5\, \exp(-s^2/2) \tag{A.25}$$

Thus for n sufficiently large

$$1 - \tilde{G}_n(b\, \sqrt{\tfrac{k_2 \log(n)}{n}}) = \exp(-\frac{k_2 b^2 \log(n)}{2v_0^2} + o(n^0))) = O(\frac{\log(n)^{5/2}}{n^{1+\delta}}) \tag{A.26}$$

for some $\delta > 0$. The same goes for $\tilde{G}_n(-2b\sqrt{\frac{\log(n)}{n}})$, and therefore, $R_n = \mathrm{O}(\log(n)^{7/2}/n^{1+\delta}) = \mathrm{o}(n^{-1})$ and

$$n\,\mathrm{MSE}(S_n, Q_n \,\big|\, D_{k,\tilde{t}}) = (n^{\natural})^{-2} \int_{-bl_n}^{bl_n} \frac{u^2}{\sqrt{n}}\; G_n'(\frac{u}{\sqrt{n}})\, du + \mathrm{o}(n^{-1}) \tag{A.27}$$

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 $t^{\natural} := \frac{t}{\sqrt{n}}, \tilde{s}_{n,k}(x) = s_{n,k}(\frac{x}{\sqrt{n}})$. 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 (A.27) more explicitly. Denote

$$\mathcal{G}_n(s,t) := G_{n,t}(s), \quad G_{n,t}^{(1)}(s) := [\tfrac{\partial}{\partial s}\mathcal{G}_n](s,t), \quad G_{n,t}^{(2)}(s) := [\tfrac{\partial}{\partial t}\mathcal{G}_n](s,t) \tag{A.28}$$

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

$$\tilde{G}_n'(\tfrac{u}{\sqrt{n}}) = [G_{n,x}^{(1)}(s(x))s'(x) + G_{n,x}^{(2)}(s(x))]\Big|_{x=\frac{u}{\sqrt{n}}} = G_{n,u/\sqrt{n}}^{(1)}(\tilde{s}(u))\,\tilde{s}'(u)\sqrt{n} + G_{n,u/\sqrt{n}}^{(2)}(\tilde{s}(u)) =: \tilde{g}_n(u)\sqrt{n}$$

and therefore

$$n\,\mathrm{MSE}(S_n, Q_n \,\big|\, D_{k,\tilde{t}}) = (n^{\natural})^{-2} \int_{-bl_n}^{bl_n} u^2 \tilde{g}_n(u)\, du + \mathrm{o}(n^{-1}) \tag{A.29}$$

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`

$$\begin{aligned}
\tilde{s}(u) \;&=\; \frac{-t^{\natural} - \sqrt{n}L(\frac{u}{\sqrt{n}})}{V(\frac{u}{\sqrt{n}})} \;=\; \tfrac{1}{v_0}\Big[(u - t^{\natural}) - \tfrac{u}{\sqrt{n}}(\tfrac{l_2 u}{2} + \tilde{v}_1(u - t^{\natural})) + \\
&\quad + \tfrac{1}{n}\Big((l_2\tfrac{\tilde{v}_1}{2} - \tfrac{l_3}{6})u^3 + (u - t^{\natural})u^2(\tilde{v}_1^2 - \tilde{v}_2/2)\Big)\Big] + \mathrm{O}(n^{-(1+\delta)}) \qquad (A.30)\\
\tilde{s}'(u) \;&=\; -\frac{L'(\frac{u}{\sqrt{n}})}{V(\frac{u}{\sqrt{n}})} + \frac{(t^{\natural} + L(\frac{u}{\sqrt{n}}))V'(\frac{u}{\sqrt{n}})}{V^2(\frac{u}{\sqrt{n}})} \;=\; \tfrac{1}{v_0}\Big[1 - l_2\tfrac{u}{\sqrt{n}} - 2\tilde{v}_1\tfrac{u}{\sqrt{n}} + \tfrac{t^{\natural}}{\sqrt{n}}\tilde{v}_1 + \\
&\quad + \tfrac{1}{n}\Big((3\tilde{v}_1^2 - \tfrac{l_3}{2} - \tfrac{3}{2}\tilde{v}_2 + \tfrac{3}{2}\tilde{v}_1 l_2)u^2 + ut^{\natural}(\tilde{v}_2 - 2\tilde{v}_1^2)\Big)\Big] + \mathrm{O}(n^{-(1+\delta)}) \qquad (A.31)
\end{aligned}$$

as well as

$$\begin{aligned}
G_{n,u/\sqrt{n}}^{(1)}(\tilde{s}) \;&=\; \varphi(\tilde{s})\Big[1 + \tfrac{1}{6\sqrt{n}}(\rho_0 + \rho_1\tfrac{u}{\sqrt{n}})(\tilde{s}^3 - 3\tilde{s}) + \tfrac{1}{24n}\kappa_0(\tilde{s}^4 - 6\tilde{s}^2 + 3) + \\
&\quad + \tfrac{1}{72n}\rho_0^2(\tilde{s}^6 - 15\tilde{s}^4 + 45\tilde{s}^2 - 15)\Big] + \mathrm{O}(n^{-(1+\delta)}) \qquad (A.32)
\end{aligned}$$

and respectively, $G_{n,u/\sqrt{n}}^{(2)}(\tilde{s}) = \varphi(\tilde{s})\frac{\rho_1}{6\sqrt{n}}(1 - \tilde{s}^2) + \mathrm{O}(n^{-(1/2+\delta)})$. This gives

$$\tilde{g}_n(u) = v_0\varphi(\tilde{s})[1 + \tfrac{1}{\sqrt{n}}P_1(u, t^{\natural}) + \tfrac{1}{n}P_2(u, t^{\natural})] + \mathrm{O}(n^{-(1+\delta)}) \tag{A.33}$$

for

$$P_1(u, t^{\natural}) = -l_2 u - 2\tilde{v}_1 u + t^{\natural}\tilde{v}_1 + \tfrac{\rho_0}{6v_0^3}(u - t^{\natural})^3 - \tfrac{\rho_0}{2v_0}(u - t^{\natural}) \tag{A.34}$$

and $P_2(u, t^{\natural})$ a corresponding polynomial in $u, t^{\natural}, \tilde{v}_1, \tilde{v}_2, l_2, l_3, \rho_0, \rho_1$, 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 $s_1 = (u - t^{\natural})/v_0$ as

$$\varphi(\tilde{s}) = \varphi(s_1)[1 - s_1(\tilde{s} - s_1) + (s_1^2 - 1)(\tilde{s} - s_1)^2/2] + \mathrm{O}(n^{-(1+\delta)}) \tag{A.35}$$

and hence $\tilde{g}_n(u) = v_0\varphi(s_1)g_n(s_1) + \mathrm{O}(n^{-(1+\delta)})$ with $g_n(s_1) := 1 + \frac{1}{\sqrt{n}}\tilde{P}_1(s_1, t^{\natural}) + \frac{1}{n}\tilde{P}_2(s_1, t^{\natural})$ for

$$\tilde{P}_1(s_1, t^{\natural}) = \rho_0\frac{s_1^3 - 3s_1}{6} + (\tfrac{l_2}{2} + \tilde{v}_1)s_1^3 - (l_2 + 2\tilde{v}_1)s_1 v_0 + (l_2 + \tilde{v}_1)[s_1^2 - 1]t^{\natural} + \tfrac{(t^{\natural})^2 l_2 s_1}{2v_0} \tag{A.36}$$

and $\tilde{P}_2(s_1, t^\natural)$ a corresponding polynomial again to be looked up from our MAPLE procedure `asgns`. This gives

$$n\,\mathrm{MSE}(S_n, Q_n \,\big|\, D_{k,\tilde{t}}\,) = (n^\natural)^{-2} \int_{-bl_n/v_0}^{bl_n/v_0} h_n(s)\varphi(s)\,\lambda(ds) + \mathrm{o}(n^{-1}) \tag{A.37}$$

for

$$h_n(s) = u_1(s)^2 g_n(s), \qquad u_1(s) = s v_0 + t^\natural \tag{A.38}$$

Selection of the least favorable contamination Function $h_n(s)$ from (A.38) 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

$$h_n(s) = (sv_0 + t^\natural)^2 + \frac{1}{\sqrt{n}} Q_1 + \frac{1}{n} Q_2 \tag{A.39}$$

where $v_0 Q_1$ 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 $v_0^2 Q_2$ 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 $h_{n,t,t}(s) = 2 + \frac{1}{\sqrt{n}} Q_{1,t,t} + \frac{1}{n} Q_{2,t,t}$ where $\deg(Q_{1,t,t}, s) = 3$ and $\deg(Q_{2,t,t}, s) = 6$, and under symmetry, more specifically

$$l_2 = \tilde{v}_1 = \rho_0 = 0 \tag{A.40}$$

$Q_{1,t,t} = 0$ 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 (A.40), 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, $t^\natural = \tilde{t}/\sqrt{\bar{n}} = \sum_{i:U_i=1} \psi(X_i - t)/\sqrt{\bar{n}}$. 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 A.3.

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 (3.12) resp. (3.13): On (I), $|t^\natural|$ is bounded, so smallness of the probabilities in (3.12) resp. (3.13) 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 (3.12) resp. (3.13) 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

$$\tilde{k} := k/\sqrt{n}, \qquad k^\natural := k/\sqrt{\bar{n}} = \tilde{k}/n^\natural \tag{A.41}$$

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

$$\tilde{h}_n(s) = \tilde{h}_n(s, k^\natural) = h_n(s, k^\natural b) \tag{A.42}$$

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 B.4, we may drop the integration limits and get

$$n\,\mathrm{MSE}(S_n, Q_n^0 \,\big|\, K = k\,) = (n^\natural)^{-2} \int_{-\infty}^{\infty} \tilde{h}_n(s)\varphi(s)\,\lambda(ds) + \mathrm{o}(n^{-1}) \tag{A.43}$$

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

$$\mathrm{E}[X^2] = 1, \quad \mathrm{E}[X^4] = 3, \quad \mathrm{E}[X^6] = 15, \quad \mathrm{E}[X^8] = 115 \tag{A.44}$$

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

$$\begin{aligned} n\,\mathrm{MSE}(S_n, Q_n^0 \,\big|\, K = k\,) \;=\; & \mathrm{o}(n^{-1}) + (n^\natural)^{-2}\Big[(k^\natural)^2 b^2 + v_0^2 + \tfrac{1}{\sqrt{n}}[\pm(3l_2 + 4\tilde{v}_1)v_0^2 k^\natural b \pm l_2 (k^\natural)^3 b^3] + \\ & + \tfrac{1}{n}[(\tfrac{5}{4} l_2^2 + \tfrac{1}{3} l_3)(k^\natural)^4 b^4 + (3\tilde{v}_2 + 2l_3 + 3\tilde{v}_1^2 + \tfrac{15}{2} l_2 + 12\tilde{v}_1 l_2) v_0^2 (k^\natural)^2 b^2 + \\ & + (\rho_0(2\tilde{v}_1 + l_2) + \tfrac{2}{3}\rho_1) v_0^3) + (12\tilde{v}_1 l_2 + l_3 + 3\tilde{v}_2 + \tfrac{15}{4} l_2^2 + 9\tilde{v}_1^2) v_0^4]\Big] \end{aligned} \tag{A.45}$$

As mentioned in Remark 3.6(c), the terms of κ_0 cancel out for A_2 as do the terms of ρ_0 for A_1.

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

$$(n^\natural)^{-2} = 1 + \frac{\tilde k}{\sqrt{n}} + \frac{\tilde k^2}{n} + \mathrm{o}(n^{-1}), \qquad (n^\natural)^{-3} = 1 + \frac{3\tilde k}{2\sqrt{n}} + \mathrm{o}(n^{-1/2}), \qquad (n^\natural)^{-4} = 1 + \mathrm{o}(n^0) \tag{A.46}$$

$$k^\natural = \tilde k + \frac{\tilde k^2}{2\sqrt{n}} + \mathrm{o}(n^{-1/2}),\ \ (k^\natural)^2 = \tilde k^2 + \frac{\tilde k^3}{\sqrt{n}} + \frac{\tilde k^4}{n} + \mathrm{o}(n^{-1}),\ \ (k^\natural)^3 = \tilde k^3 + \frac{3\tilde k^4}{\sqrt{n}} + \mathrm{o}(n^{-1/2}),\ \ (k^\natural)^4 = \tilde k^4 + \mathrm{o}(n^0) \tag{A.47}$$

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

$$\begin{aligned} n\,\mathrm{MSE}(S_n, Q_n^0\,|\,K=k) &= \mathrm{o}(n^{-1}) + \tilde k^2 b^2 + v_0{}^2 + \frac{[\pm(4\,\tilde v_1 + 3\,l_2\,)\,b+1]\tilde k v_0{}^2 + (2 \pm l_2 b)\,\tilde k^3 b^2}{\sqrt{n}} + \\ &+ \frac{\Big(3\,b^2 \pm 3\,l_2\,b^3 + \big(\tfrac54\,l_2^2 + \tfrac13\,l_3\big)\,b^4\Big)\tilde k^4 + \big(3\tilde v_2 + 9\tilde v_1^2 + \tfrac{15}{4}\,l_2^2 + l_3 + 12\,l_2\,\tilde v_1\big)\,v_0{}^4 + \big((l_2 + 2\tilde v_1)\,\rho_0 + \tfrac23\,\rho_1\big)\,v_0{}^3}{n} + \\ &+ \frac{\Big(\big(3\,\tilde v_1^2 + 3\,\tilde v_2 + 12\,l_2\tilde v_1 + \tfrac{15}{2}\,l_2^2 + 2\,l_3\big)\,b^2 + 1 \pm (6\,l_2\, + 8\,\tilde v_1\,)\,b\Big)\tilde k^2 v_0^2}{n} \end{aligned} \tag{A.48}$$

Integration w.r.t. $\tilde k$ As by Corollary A.2 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 A.1 to determine the MSE. This gives the result by our `MAPLE` procedures `intekout`, `asMSE`:

$$n\ \mathrm{E}_{Q_n^0}[S_n^2] = r^2 b^2 + v_0{}^2 + \frac{r}{\sqrt{n}}\,A_1 + \frac1n\,A_2 + \mathrm{o}(n^{-1}) \tag{A.49}$$

with

$$A_1 \;=\; v_0{}^2\,\Big(\pm(4\,\tilde v_1 + 3\,l_2\,)b + 1\Big) + b^2 + [2\,b^2 \pm l_2\,b^3\,]\,r^2 \tag{A.50}$$

$$\begin{aligned} A_2 \;=\; & v_0{}^3\,\Big((l_2 + 2\,\tilde v_1\,)\rho_0 + \tfrac23\,\rho_1\Big) + v_0{}^4\,(3\,\tilde v_2 + \tfrac{15}{4}\,l_2^2 + l_3 + 9\,\tilde v_1^2 + 12\,\tilde v_1\,l_2\,) + \\ & +[\,v_0{}^2\,\Big((3\,\tilde v_2 + 3\,\tilde v_1^2 + \tfrac{15}{2}\,l_2^2 + 2\,l_3 + 12\,\tilde v_1\,l_2\,)b^2 + 1 \pm (8\,\tilde v_1 + 6\,l_2\,)\,b\Big) \pm 3\,l_2\,b^3 + 5\,b^2\,]\,r^2 + \\ & +\Big((\tfrac54\,l_2^2 + \tfrac13\,l_3\,)b^4 \pm 3\,l_2\,b^3 + 3\,b^2\Big)\,r^4 \end{aligned} \tag{A.51}$$

Decision upon the alternative (3.12) *or* (3.13) Denote Q_n^- a contaminated member in $\tilde{\mathcal{Q}}_n(r)$ according to (3.12) and correspondingly Q_n^+ according to (3.13). With respect to terms of (A.49)–(A.51), 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:

$$n(\mathrm{E}_{Q_n^+}[S_n^2] - \mathrm{E}_{Q_n^-}[S_n^2]) = \frac{rb}{\sqrt{n}}\Big\{l_2[(r^2b^2 + 3v_0^2)(1 + 2\frac{r}{\sqrt{n}}) + \frac{3b^2r(r^2+1)}{\sqrt{n}}] + 4v_0^2(1 + \frac{r}{\sqrt{n}})v_1\Big\} + \mathrm{o}(n^{-1}) \tag{A.52}$$

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

$$\tilde v_1 \;>\; [<] - \frac{l_2}{4}\Big(\frac{b^2}{v_0^2}(r^2+3)(1 + \frac{r}{\sqrt{n}} - \frac{2r^2}{n}) + 3(1 - \frac{b^2}{v_0^2})\Big) \tag{A.53}$$

If there is “=” in (A.53), no decision can be taken up to order $\mathrm{o}(n^{-1})$. □

A.4 Proofs to Propositions 6.1 and 6.2

For $\varepsilon_1 \in (0,1)$, let $N_+(t) = N_+(t; n, \varepsilon_1, \hat b)$, $N_-(t) = N_-(t; n, \varepsilon_1, \check b)$ be defined as

$$N_+(t) := \#\Big\{\psi(x_i - t) \ge \hat b(1-\varepsilon_1),\ U_i = 0\Big\}, \qquad N_-(t) := \#\Big\{\psi(x_i - t) \le \check b(1-\varepsilon_1),\ U_i = 0\Big\} \tag{A.54}$$

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

$$\{\textstyle\sum\psi(x_i - t) \le 0\} \subset \{N_+(t) \le n_+\},\ \ \{\textstyle\sum\psi(x_i - t) \ge 0\} \subset \{N_-(t) \le n_-\} \tag{A.55}$$

for some numbers n_-, n_+ yet to be specified.
For Proposition 6.1, symbolically in the tableau of page 19, 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 (A.55). In this case, $\{\sum \psi(x_i - t) \le 0\}$ 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

$$\sum \psi(x_i - t) \le 0 \qquad \Longrightarrow \qquad N_+\hat{b}(1 - \varepsilon_1) + K\hat{b} + N''\check{b} \le 0 \tag{A.56}$$

that is $N_+ \le \Big(- n\check{b} - K(\hat{b} - \check{b})\Big)\Big/\Big(\hat{b}(1 - \varepsilon_1) - \check{b}\Big)$, and as this has to hold for all $K \le \varepsilon_0' n$, $N_+ \le n\Big(- \check{b} - \varepsilon_0'(\hat{b} - \check{b})\Big)\Big/\Big(\hat{b}(1 - \varepsilon_1) - \check{b}\Big) =: n_+ = n_+(\varepsilon_0')$, where by (6.2) and as $0 < \varepsilon_1 < 1$, we get $n_+ = n\varepsilon_+$ for

$$0 < \varepsilon_+ = \Big(- \check{b} - \varepsilon_0'(\hat{b} - \check{b})\Big)\Big/\Big(\hat{b}(1 - \varepsilon_1) - \check{b}\Big) < 1 - \varepsilon_0' \tag{A.57}$$

Accordingly, for the second inclusion in (A.55), we obtain

$$N_- \le n\varepsilon_- =: n_- = n_-(\varepsilon_0') \qquad \text{for} \quad \varepsilon_- := \Big(\hat{b} - \varepsilon_0'(\hat{b} - \check{b})\Big)\Big/\Big(\hat{b} - \check{b}(1 - \varepsilon_1)\Big) \tag{A.58}$$

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

$$\Pr\{S_n > \sqrt{t} \,\Big|\, D_{k,\tilde{t}=-k\check{b}}\} \overset{(1.9)}{\le} \Pr\Big\{T_{n,k}(\sqrt{t}\,) \ge k\check{b}\Big\} \le \Pr\Big\{T_{n,k}(\sqrt{t}\,) \ge \bar{k}\check{b}\Big\} \le \Pr\Big\{N_-(\sqrt{t}\,) \le n_- \,\Big|\, K = \bar{k}\Big\} \tag{A.59}$$

and correspondingly $\Pr\Big\{S_n < -\sqrt{t} \,\Big|\, D_{k,\tilde{t}=k\hat{b}}\Big\} \le \Pr\Big\{N_+(-\sqrt{t}\,) \le n_+ \Big|\, K = \bar{k}\Big\}$. But, $\mathcal{L}(N_\pm | K = k)$ is $\mathrm{Bin}(n - k, p_\pm)$ for

$$p_-(t) = \Pr\big(\psi(X^{\mathrm{id}} - \sqrt{t}\,) \le \check{b}(1 - \varepsilon_1)\big) =, \qquad p_+(t) = \Pr\big(\psi(X^{\mathrm{id}} + \sqrt{t}\,) \ge \hat{b}(1 - \varepsilon_1)\big) \tag{A.60}$$

That is, $p_-(t) = F(\sqrt{t} + B_-)$, $p_+(t) = \bar{F}(\sqrt{t} - B_+)$ where $\bar{F} = 1 - F$ and

$$B_- := \inf\Big\{y \,\Big|\, \psi(y) \ge (1 - \varepsilon_1)\check{b}\Big\}, \qquad B_+ := \sup\Big\{y \,\Big|\, \psi(y) \le (1 - \varepsilon_1)\hat{b}\Big\} \tag{A.61}$$

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 (A.59), we obtain $\binom{m}{j} \le 2^n$, $j = 0, \ldots m_\pm$, and $p_-(t), (1 - p_+(t)) \le 1$, so that

$$\sup_k \Pr\Big\{|S_n| > \sqrt{t} \,\Big|\, D_{k,|\tilde{t}|=k\hat{b}}\Big\} \le n2^n p_t^{[m-(m_- \vee m_+)]} \tag{A.62}$$

But by (A.57), $1 - \varepsilon_0' - (\varepsilon_- \vee \varepsilon_+) =: \alpha > 0$, so $m - (m_- \vee m_+') \ge \alpha n - 1$. 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 (A.55). Thus

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

Let $\tilde{p}_+ = \bar{F}(\sqrt{t} + B_+)$. To $\delta > 0$ there is a $T > 0$ 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 $t > T'$

$$t^{1/n'}(1 - F(t)) > c \qquad \Longleftrightarrow \qquad (1 - F(t))^{n'} > c^{n'} t^{-1} \tag{A.63}$$

Then for the M-estimator S_n,

$$\mathrm{E}_F[(\mathrm{S}_n)_+^2] \ge \int_{(T')^2}^\infty \Pr\Big\{S_n > \sqrt{t}\,\Big\}\, dt \ge \int_{(T')^2}^\infty \binom{n}{n'}(1 - \delta)c^{n'}(\sqrt{t} + B_+)^{-1}\, dt = \infty \tag{A.64}$$

□

A.5 Proof of Proposition 6.3

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

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

Let $A_{k,t} := \left\{\sum_{i:\,U_i=1} \psi(X_i - \sqrt{t}\,) \le (k-1)(b - c_0/\sqrt{n}\,)\right\}$. Hence if $t > v_0^2 \log(n)/n$, by (6.6), for all $k > (1-\delta)r\sqrt{n}$,

$$\Pr(A_{k,t} \,\Big|\, K = k) \ge p_0 \tag{A.65}$$

Now we proceed as in section A.3, and even with restriction (A.65) the arguments of subsection A.3 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 (A.37). 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 A.3, we obtain for $\varDelta_n = n\Big(\mathrm{MSE}(S_n, Q_n^0 \,\big|K = k\,) - \mathrm{MSE}(S_n, Q_n^\flat \,\big|K = k\,)\Big)$

$$\varDelta_n \ge p_0 \int_{\sqrt{\log n}}^{bl_n/v_0} \Big(2v_0 s D_n(\tilde{k}) + 2\tilde{k} b D_n(\tilde{k}) - D_n(\tilde{k})^2\Big)\varphi(s)\, ds + \mathrm{o}(n^{-1})$$

for $D_n(\tilde{k}) = \tilde{k}c_0/\sqrt{n} + b/\sqrt{n} + \mathrm{o}(1/\sqrt{n})$. 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 $\varphi(a_2) = \mathrm{o}(n^{-1})$,

$$\varDelta_n \ge \frac{p_0}{\sqrt{2\pi n}}[2v_0 D_n(\tilde{k}) - 2\tfrac{\tilde{k}bD_n(\tilde{k}) + D_n(\tilde{k})^2}{bl_n/v_0}] + \mathrm{o}(n^{-1}) = \tfrac{2p_0 v_0}{\sqrt{2\pi n}} D_n(\tilde{k}) + \mathrm{o}(n^{-1})$$

Now the restriction to $(1-\delta)r\sqrt{n} < K < k_1 r\sqrt{n}$ by Lemma A.1 may be dropped, giving $\varDelta_n \ge \frac{2p_0 v_0}{n\sqrt{2\pi}}(rc_0 + b) + \mathrm{o}(n^{-1})$. □

B Auxiliary Results

B.1 Two Hoeffding Bounds

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

$$P(\frac{1}{n}\sum_i \xi_i - \mathrm{E}[\xi_1] \ge \varepsilon) \le \exp(-\frac{2n\varepsilon^2}{M^2}), \qquad P(\frac{1}{n}\sum_i \xi_i - \mathrm{E}[\xi_1] \le -\varepsilon) \le \exp(-\frac{2n\varepsilon^2}{M^2}) \tag{B.1}$$

Proof Hoeffding [12, Thm. 2. and Thm. 1, inequality (2.1)]. □

B.2 A uniform Edgeworth expansion

In the following theorem, generalizes Ibragimov [15, Thm. 1] and Ibragimov and Linnik [16, Thm. 3.3.1] to the situation where the law of ξ_i depends through an additional parameter t:

Theorem B.2 *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*

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

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

$$F_n(s,t) := P(\textstyle\sum_{i=1}^n \xi_{i,t} < s\sqrt{n}), \qquad H_n(s,t) := \Phi(s) - \dfrac{\varphi(s)}{\sqrt{n}}\varphi(s)\dfrac{\rho_t}{6}(s^2 - 1) \tag{B.3}$$

$$G_n(s,t) := H_n(s,t) - \frac{\varphi(s)}{n}\Big[\frac{\kappa_t}{24}(s^3 - 3s) + \frac{\rho_t^2}{72}(s^5 - 10s^3 + 15s)\Big] \tag{B.4}$$

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)'*

$$\hat{f}_{u_0}(u_1) := \sup_{u_0<u<u_1} \sup_t |f_t(u)| < 1 \tag{B.5}$$

is fulfilled, then

$$\sup_{s\in\mathbb{R}} \sup_t |F_n(s,t) - H_n(s,t)| = \mathrm{o}(n^{-1/2}) \tag{B.6}$$

(b) *If* $\sup_t \mathrm{E}\,|\xi_{i,t}|^5 < \infty$ *and the uniform Cramér–condition (C)*

$$\lim_{u\to\infty} \sup_t |f_t(u)| < 1 \tag{B.7}$$

is fulfilled, then

$$\sup_{s\in\mathbb{R}} \sup_t |F_n(s,t) - G_n(s,t)| = \mathrm{O}(n^{-3/2}) \tag{B.8}$$

Proof The general technique to prove Edgeworth expansions is to use Berry's smoothing lemma, which we take from Ibragimov and Linnik [16, 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 $G_{n,1,t}(s) = H_n(s,t)$, $G_{n,2,t}(s) = G_n(s,t)$ as well as their Fourier-Stieltjes transforms $g_{n,j,t}(u) = \int e^{isu} G'_{n,j,t}(s)\,\lambda(ds)$ and $\hat{G}'_{n,j} = \sup_t \sup_{s\in\mathbb{R}} |G'_{n,j,t}(s)|$. Then for $T > T' > 0$, it holds that

$$\begin{aligned}\sup_{s\in\mathbb{R}} \sup_t |F_n(s,t) - G_{n,j,t}(s)| \le \sup_t \frac{1}{\pi} \int_{-T'}^{T'} \frac{|f_{n,t}(u) - g_{n,j,t}(u)|}{|u|}\,\lambda(du) + \sup_t \frac{1}{\pi} \int_{T'\le|u|<T} \frac{|f_{n,t}(u)|}{|u|}\,\lambda(du) + \\ + \sup_t \frac{1}{\pi} \int_{T'\le|u|<T} \frac{|g_{n,j,t}(u)|}{|u|}\,\lambda(du) + \sup_t \frac{24}{\pi T} \hat{G}'_{n,j}\end{aligned} \tag{B.9}$$

But similarly as in Ibragimov [15, pp. 462], for some constants $\gamma > 0$ and $c_j > 0$, we get for $T' = \gamma\sqrt{n}$ and $|u| \le T'$

$$\frac{|f_{n,t}(u) - g_{n,j,t}(u)|}{|u|} \le c_j \sup_t \mathrm{E}[|\xi_{1,t}|^{3+j}]\, n^{-(j+1)/2}\,(|u|^j + |u|^{2+3j})\, e^{-u^2/4} \tag{B.10}$$

and hence the first summand in the RHS of (B.9) 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

$$\int_{T'}^{T} \frac{|f_{n,t}(u)|}{u}\,\lambda(du) = \int_{\gamma}^{T/\sqrt{n}} \frac{|f_t^n(u)|}{u}\,\lambda(du) \tag{B.11}$$

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

$$\sup_t \int_{T'}^{T} \frac{|f_{n,t}(u)|}{u}\,\lambda(du) \le \log(T/\sqrt{n}\,)\beta^n = \mathrm{o}(e^{-\sqrt{n}/2}) \tag{B.12}$$

In case $j = 1$, we proceed as in Ibragimov and Linnik [16, Lemma 3.3.1]: If $\sup_{u_1} \hat{f}_\gamma(u_1) < 1$ 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 in u_1 and tends to 1. So we may define $l'_n := \inf\{u_1 \,\big|\, \hat{f}_\gamma(u_1) \ge 1 - 1/\sqrt{n}\,\}$. 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

$$\sup_t \int_{T'}^{T} \frac{|f_{n,t}(u)|}{u}\,\lambda(du) \le \log(\sqrt{n})(1 - 1/\sqrt{n})^n \le \log(\sqrt{n})e^{-\sqrt{n}} = \mathrm{o}(e^{-\sqrt{n}/2}) \tag{B.13}$$

Hence the second summand in the the RHS of (B.9) 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 (B.9) 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 $\mathrm{O}(\exp(-\gamma^2 n/3))$. □

B.3 Moments for the Binomial

Lemma B.3 *Let* $X \sim \mathrm{Bin}(n, p)$*. Then*

$$\mathrm{E}[X] = pn, \qquad \mathrm{E}[X^2] = p^2n^2 + pn - p^2n, \tag{B.14}$$

$$\mathrm{E}[X^3] = p^3n^3 - 3p^3n^2 + 2p^3n + 3p^2n^2 - 3p^2n + pn, \tag{B.15}$$

$$\mathrm{E}[X^4] = p^4n^4 - 6p^4n^3 + 11p^4n^2 - 6p^4n + 6p^3n^3 - 18p^3n^2 + 12p^3n + 7p^2n^2 - 7p^2n + pn \tag{B.16}$$

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

$$\mathrm{E}[X] = rn^{1/2}, \qquad \mathrm{E}[X^2] = r^2n + rn^{1/2} - r^2, \tag{B.17}$$

$$\mathrm{E}[X^3] = r^3n^{3/2} + 3r^2n + (r - 3r^3)n^{1/2} - 3r^2 + 2r^3n^{-1/2}, \tag{B.18}$$

$$\mathrm{E}[X^4] = r^4n^2 + 6r^3n^{3/2} + (7r^2 - 6r^4)n + (r - 18r^3)n^{1/2} + 11r^4 - 7r^2 + 12r^3n^{-1/2} - 6r^4n^{-1} \tag{B.19}$$

Proof easy calculations for `MAPLE` — see procedure `Binmoment`... □

B.4 Decay of the standard normal

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

Lemma B.4 *Let* $X \sim \mathcal{N}(0, 1)$*. Then for* $0 \le k \le 8$ *and any sequence* $(c_n)_n \subset \mathbb{R}$ *with* $\liminf_n c_n > \sqrt{2}$*,*

$$\mathrm{E}[X^k\, \mathrm{I}_{\{X \ge c_n \sqrt{\log(n)}\}}] = \mathrm{o}(n^{-1}) \tag{B.20}$$

Proof Ruckdeschel [25, Lem. A.6]. □

References

1. 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.
2. Barndorff-Nielsen O. and Cox D. [1994]: *Inference and asymptotics*, Vol. 52 of *Monographs on Statistics and Applied Probability*. Chapman and Hall.
3. Bhattacharya R. and Rao R. [1976]: *Normal approximation and asymptotic expansions*. John Wiley&Sons, Inc.
4. Bickel P.J., Klaassen C.A., Ritov Y. and Wellner J.A. [1998]: *Efficient and adaptive estimation for semiparametric models*. Springer.
5. Daniels H.E. [1954]: Saddlepoint approximations in statistics. *Ann. Math. Statistics*, **25**: 631–650.
6. Donoho D.L. and Huber P.J. [1983]: The notion of breakdown point. In: Bickel, P.J., Doksum, K.A. and Hodges, J.L.jun. (eds.) *Festschr. for Erich L. Lehmann*, pp. 157–184.
7. Field C. and Ronchetti E. [1990]: *Small sample asymptotics*, Vol. 13 of *IMS Lecture Notes - Monograph Series.*. Institute of Mathematical Statistics, Hayward, CA.
8. Fraiman R., Yohai V.J. and Zamar R.H. [2001]: Optimal robust M-estimates of location. *Ann. Stat.*, **29**(1): 194–223.
9. Gusev S. [1976]: Asymptotic expansions associated with some statistical estimators in the smooth case. II: Expansions of moments and distributions. *Theory Probab. Appl.*, **21**: 14–33.
10. Hall P. [1992]: *The bootstrap and Edgeworth expansion*. Springer Series in Statistics. Springer-Verlag.
11. Hampel F.R. [1974]: Some small sample asymptotics. In: *Proc. Prague Symp. Asympt. Stat., Vol. II, Prague 1973*, p. 109–126.
12. Hoeffding W. [1963]: Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.*, **58**: 13–30.
13. Huber P.J. [1964]: Robust estimation of a location parameter. *Ann. Math. Stat.*, **35**: 73–101.
14. —— [1981]: *Robust statistics*. Wiley.

15. Ibragimov I. [1967]: The Chebyshev-Cramér asymptotic expansions. *Theor. Probab. Appl.*, **12**: 454–469.
16. Ibragimov I. and Linnik Y. [1971]: *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing Company, Groningen. edited by J.F.C. Kingman.
17. Kohl M., Ruckdeschel P. and Rieder H. [2010]: Infinitesimally robust estimation in general smoothly parametrized models. *Statistical Methods and Applications*. To appear; DOI: 10.1007/s10260-010-0133-0.
18. Le Cam L. [1986]: *Asymptotic methods in statistical decision theory*. Springer.
19. Pfaff T. [1977]: *Existenz und asymptotische Entwicklungen der Momente mehrdimensionaler maximum likelihood-Schätzer.* Inaugural-dissertation, Universität zu Köln.
20. Pfanzagl J. [1985]: *Asymptotic expansions for general statistical models. With the assist. of W. Wefelmeyer*, Vol. 31 of *Lecture Notes in Statistics*. Springe.
21. R Development Core Team [2010]: *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*
22. Rieder H. [1994]: *Robust asymptotic statistics*. Springer Series in Statistics. Springer.
23. Rousseeuw P.J. [1984]: Least median of squares regression. *J. Am. Stat. Assoc.*, **79**: 871–880.
24. Ruckdeschel P. [2006]: A motivation For $1/\sqrt{n}$-Shrinking-Neighborhoods? *Metrika*, **63**(3): 295–307.
25. —— [2010]: Higher Order Asymptotics for the MSE of the Median on Shrinking Neighborhoods. Submitted in parallel; available on the web-page to this article.
26. Ruckdeschel P. and Kohl M. [2010a]: Computation of the Finite Sample Risk of M-Estimators on Neighborhoods. Submitted in parallel; available on the web-page to this article.
27. —— [2010b]: General Purpose Convolution Algorithm for Distributions in S4-Classes by means of FFT. Submitted in parallel; available on the web-page to this article.
28. Ruckdeschel P., Kohl M., Stabla T. and Camphausen F. [2006]: S4 classes for distributions. *R News*, **6**(2): 2–6. **URL:** *http://CRAN.R-project.org/doc/Rnews/*
29. —— [2010]: S4 *Classes for Distributions—a manual for packages* `distr`, `distrEx`, `distrEllipse`, `distrMod`, `distrSim`, `distrTEst`, *and* `distrTeach`, *version 2.3*.
URL: *http://cran.r-project.org/web/packages/distrDoc/vignettes/distr.pdf*
30. Taniguchi M. and Kakizawa Y. [2000]: *Asymptotic theory of statistical inference for time series*. Springer.
31. van der Vaart A. [1998]: *Asymptotic statistics*, Vol. 3 of *Cambridge Series on Statistical and Probabilistic Mathematics*. Cambridge Univ. Press.
32. Yohai V.J. [1987]: High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.*, **15**(2): 642–656.

Web-page to this article:
`http://www.mathematik.uni-kl.de/~ruckdesc/`