Computation of Finite Sample Risks of M-Estimators on Neighborhoods

Peter Ruckdeschel[a,b], Matthias Kohl[c]

[a] *Fraunhofer ITWM, Abt. Finanzmathematik, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany*
[b] *TU Kaiserslautern, AG Statistik, FB. Mathematik, P.O.Box 3049, 67653 Kaiserslautern, Germany*
[c] *Paul-Martini-Clinical Sepsis Research Unit, Dept. of Anesthesiology and Intensive Care Medicine, Jena University Hospital, Friedrich-Schiller-University, Erlanger Allee 101, 07747 Jena, Germany*

Abstract

We compare various approaches for the determination of finite-sample risks of one-dimensional location M-estimators on convex contamination and total variation neighborhoods. As risks we consider mean squared error (MSE) and certain over-/undershooting probabilities as in Huber (1968) and Rieder (1980). Our comparison consists of (numerically) exact formulae (via FFT, Ruckdeschel and Kohl (2010)), Edgeworth expansions, saddlepoint approximations as in Field and Ronchetti (1990), an approach by Fraiman et al. (2001), first-, second- and third order asymptotics as well as Monte-Carlo simulations.

Keywords: one-dimensional location, finite-sample risk, FFT, Edgeworth expansions, saddlepoint approximations, higher order asymptotics
2000 MSC: 93E11,62F35

1. Introduction

1.1. Motivation

Only in exceptional cases, the exact value (up to numerical errors) of the maximal finite-sample risk of a (robust) estimator on a typical robust neighborhood is accessible. This is why Robust Statistics usually recurs to asymptotics, possibly refined to higher order asymptotics, to simulations, and to bootstrapping and other resampling techniques. There are such exceptional cases, though, and it is worthwhile studying them in order to get a more precise idea of the finite-sample error incurred by any of the mentioned ways out.

The most prominent exception are M-estimators of (one-dimensional) location to a monotone ψ-function as already studied in Huber (1964), where we have (more or less) explicit formulae for their distribution in both ideal, and, for certain types of neighborhoods, i.e. convex contamination (gross error) and total variation, also in a least favorable

Email addresses: Peter.Ruckdeschel@itwm.fraunhofer.de (Peter Ruckdeschel), Matthias.Kohl@stamats.de (Matthias Kohl)

situation. Hence, we can compute corresponding risks like maximal MSE or the over-/undershooting risk of Huber (1968). More specifically these M-estimators also comprise the corresponding minimax procedure minimizing the maximal risk.

Although very simple in structure, the corresponding distributions are computationally somewhat demanding as they have both non-trivial (absolutely) continuous and discrete parts, and hence require careful treatment. In this paper, we discuss some implementations to R of these distributions as well as functions to compute the respective risks.

With these at hand, we can make quite precise statements about the errors involved in first to third order asymptotics. To get the same precision by means of crude Monte-Carlo simulation we would need billions of replications. Of course, more refined simulation techniques could reduce the respective number of simulations, but still the numerical approach is conceptually simpler as it does not need to take into account special features of the (ideal) distribution.

1.2. Other approaches

1.3. Organization of the paper

2. Setup

2.1. Ideal model

We consider one-dimensional location, i.e. for sample size n we have i.i.d. observations y_i $(i = 1, \ldots, n)$ composed by errors u_i and an unknown location parameter $\theta \in \mathbb{R}$, i.e. $y_i = \theta + u_i$. For simplicity, our reference ideal model distribution will be Gaussian, that is $u_i \overset{\text{i.i.d.}}{\sim} F = \mathcal{N}(0, 1)$. However, all numerical results are valid for arbitrary F with finite Fisher-information of location and with log-concave densities which is the setting of Huber (1981, Ch. 4, Ex. 5.2).

2.2. Deviations from the ideal model

Following Rieder (1980), we introduce the following types of neighborhoods of P_θ as deviations from the ideal model.

Definition 2.1. *For given numbers $\varepsilon, \delta \in [0, 1)$ with $\varepsilon + \delta \in (0, 1)$,*

$$\mathcal{U}_{cv}(\theta) = \mathcal{U}_{cv}(\theta; \varepsilon; \delta) = \left\{ Q \in \mathcal{M}_1(\mathbb{B}) \,\middle|\, Q(dy) \geq (1 - \varepsilon) P_\theta(dy) - \delta \right\} \tag{2.1}$$

Remark 2.2. **(a)** These neighborhoods include contamination ($\delta = 0$), abbreviated by subscript c in the sequel, as well as total variation neighborhoods ($\varepsilon = 0$), abbreviated by subscript v. A generic type will be indicated by subscript $*$. If $\delta = 0$, we write $\mathcal{U}_c(\theta) = \mathcal{U}_c(\theta; \varepsilon)$, and if $\varepsilon = 0$, we write $\mathcal{U}_v(\theta) = \mathcal{U}_v(\theta; \delta)$. In case of the MSE-risk introduced below, we write r for ε to be consistent with the literature.

(b) In case of asymptotic considerations we replace ε and δ by $\varepsilon_n = \varepsilon/\sqrt{n}$, resp. $\delta_n = \delta/\sqrt{n}$.

(c) Due to the translation invariance of both ideal model and neighborhoods, we may reduce ourselves to the parameter value $\theta = 0$; cf. Rieder (1994, Section 7.2.3), resp. Kohl (2005, Section 7.1.2) where this is verified for the linear regression model. □

As shown in Ruckdeschel (2010b), we have to thin out the original neighborhood if the loss function is unbounded, in order to obtain uniform convergence for the risk on the whole neighborhood. In case of the MSE-risk introduced below, we limit ourselves to $* = c$. To specify this thinning-out, we may – as usual – interpret $Q_n \in \mathcal{U}_c(0, \varepsilon_n)$ as the distribution of the random variable Y defined as

$$Y := (1 - U)Y^{\text{id}} + UY^{\text{di}} \tag{2.2}$$

for Y^{id}, U, Y^{di} stochastically independent, $Y^{\text{id}} \sim \mathcal{N}(0, 1)$, $U \sim \text{Bin}(1, \varepsilon_n)$, and $Y^{\text{di}} \sim P^{\text{di}}$ for some arbitrary $P^{\text{di}} \in \mathcal{M}_1$.[1] The balls $\mathcal{U}_c(\varepsilon; n)$ defined as $\{Q_n^n \mid Q_n \in \mathcal{U}_c(0, \varepsilon_n)\}$ are then thinned out to the sets $\tilde{\mathcal{U}}_c(\varepsilon; n)$ of

$$Q_n = \mathcal{L}\{[(1 - U_i)Y_i^{\text{id}} + U_i Y_i^{\text{di}}]_i \,\Big|\, \sum U_i < n/2\,\} \tag{2.3}$$

that is only those samples are retained where less than half the sample is contaminated; cf. Ruckdeschel (2010b, Section 2; e.g. Proposition 2.1 and Theorem 3.4).

2.3. Considered risks

We consider two types of finite-sample risks – (maximal) MSE (maxMSE) and the probability of over-/undershooting a certain bound. In the risk definitions we suppress the dependence on the size of the neighborhood and assume $\varepsilon, \delta \in [0, 1)$ respectively $0 < r < \sqrt{n}$ given. For the maximal MSE at sample size n, we obtain as risk of an estimator $S_n : \mathbb{R}^n \to \mathbb{R}$,

$$\text{maxMSE}(S_n) := \sup\, n\, \text{E}_Q\, S_n^2, \qquad Q \in \tilde{\mathcal{U}}_c(n) \tag{2.4}$$

For the over-/undershooting finite-sample risk at sample size n we define $\text{Risk}\,(S_n) := \sup_{\theta \in \mathbb{R}} \text{Risk}_\theta(S_n)$ with

$$\text{Risk}_\theta(S_n) = \sup \max\left\{Q_\theta^n(S_n > \theta + \tau),\, Q_\theta^n(S_n < \theta - \tau)\right\}, \quad Q_\theta \in \mathcal{U}_{cv}(\theta) \tag{2.5}$$

for a given constant $\tau \in (0, \infty)$.

Remark 2.3. Again, for an asymptotic version of this risk, we replace τ by some $\tau_n = \tau/\sqrt{n}$. In addition, the over-/undershooting bounds are allowed asymmetric; i.e., we have two bounds τ_n', τ_n'' of sum $2\tau_n$. For more details see Rieder (1980).

2.4. Minimax-Estimators

For both risks considered here, the minimax estimator may be realized as an M-estimator to monotone scores ψ. That is, following the notation in Huber (1981, pp. 46), for $\psi_t(x) = \psi(x - t)$, let

$$S_n' := \sup\left\{\, t \,\Big|\, \sum_{i \le n} \psi_t(x_i) > 0\right\}, \qquad S_n'' := \inf\left\{\, t \,\Big|\, \sum_{i \le n} \psi_t(x_i) < 0\right\} \tag{2.6}$$

and the minimax estimator is an equal randomization between S_n' and S_n''.

[1] id referring to *ideal*, di alluding to *disturbing*

In the Gaussian location model, for all types of risks and neighborhoods considered here, the minimax scores are of Huber form

$$\psi^{(c)}(u) = u \min\{1, c/|u|\} \tag{2.7}$$

The corresponding influence curve (IC) in sense of Rieder (1994, Definition 4.2.10) reads

$$\eta^{(c)}(u) = Au \min\{1, b/(A|u|)\} \tag{2.8}$$

where $b = Ac$.

Remark 2.4. The assertion that the minimax estimator may be realized as an M-estimator to scores of form (2.7) is drawn from several references: For risk (2.5), this is shown in Huber (1968); in a corresponding asymptotic setup the same class of estimators arises in Rieder (1980). There is also an extension to (univariate) linear regression, cf. Rieder (1989, 1995) and Kohl (2005). For the first order asymptotic maximal MSE the assertion is shown in Rieder (1994, Theorem 5.5.7) and in the construction part in Section 6.2.1 (ibid.). For the second order asymptotic maximal MSE it is shown in Ruckdeschel (2010a,b). For the third order asymptotic maximal MSE we do not know whether the assertion holds. □

2.4.1. MSE Risk

Higher Order Expansion for M-Estimators. In the location model of Section 2.1 – with F not necessarily Gaussian –, in Ruckdeschel (2010b), we obtain an expansion for $\text{maxMSE}(S_n)$ for M-estimators to monotone scores of form

$$\text{maxMSE}(S_n) = r^2 b^2 + {v_0}^2 + \tfrac{r}{\sqrt{n}}\, A_1 + \tfrac{1}{n}\, A_2 + \text{o}(n^{-1}) \tag{2.9}$$

where b is the sup-norm and v_0 the $L_2(P)$-norm of the corresponding IC and A_1 and A_2 are certain constants (in n) depending on r and the IC. The expansion up to $\text{o}(1)$ will be called first-order, up to $\text{o}(1/\sqrt{n}\,)$ second-order and up to $\text{o}(1/n)$ third-order asymptotics. Explicit expressions for A_1 and A_2 can be found in Ruckdeschel (2010b). The Gaussian case is spelt out in Appendix A.1.

Asymptotically optimal c. Given expansion 2.9 we may, depending on r, determine first-, second-, and third-order optimal clipping bounds c. Details on these bounds are spelt out in Appendix A.2.

Approach of Fraiman et al. (2001). In Fraiman et al. (2001), the authors work in a similar setup, that is, the one-dimensional location problem, where the center distribution is $F = \mathcal{N}(0, \sigma^2)$. They search for an M-estimator S_n to skew symmetric scores ψ which minimizes the maximal risk on a neighborhood about F. Contrary to our approach, the authors work with convex contamination neighborhoods $\mathcal{V} = \mathcal{V}(F, \varepsilon)$ of fixed radius ε. Some more details on this approach are gathered in Appendix A.3 To be able to compare their procedure to ours, for fixed sample size n, we translate their radius ε into our $r/\sqrt{n}$, and determine the optimal $\hat{\psi}_{a,b,c,t}(x)$ and the corresponding c^{FYZ} for the Huber-scores $\psi^{(c)}$ "next" to $\hat{\psi}_{a,b,c,t}(x)$. Anyway, $\hat{\psi}_{a,b,c,t}(x)$ and $\psi^{(c)}$ are practically indistinguishable; see Fraiman et al. (2001, simplification 2, p. 206).

2.4.2. Over-/Undershooting Risk

Contamination Case. **Finite Sample Setup:** For the minimax solution as in Huber (1968), c_c^{fi} is the unique solution to

$$\frac{\varepsilon_n}{1-\varepsilon_n} = \exp(-2c_c^{\mathrm{fi}}\tau_n)\Phi(\tau_n - c_c^{\mathrm{fi}}) - \Phi(-\tau_n - c_c^{\mathrm{fi}}) \tag{2.10}$$

Asymptotic Setup: In an asymptotic setup corresponding to this over-/undershooting risk (cf. Rieder (1980)), c_c^{as} is found as the unique solution to

$$\varepsilon = 2\tau\big[\varphi(c_c^{\mathrm{as}}) - c_c^{\mathrm{as}}\Phi(-c_c^{\mathrm{as}})\big] \tag{2.11}$$

The derivation of this result can be found in Kohl (2005, Section 10.3). Some relations between finite and asymptotic bounds are listed in Appendix A.4.

Total Variation Case. **Finite Sample Setup:** For the minimax solution as in Huber (1968), c_v^{fi} is the unique solution to

$$\big[1+\exp(-2c_v^{\mathrm{fi}}\tau_n)\big]\delta_n = \exp(-2c_v^{\mathrm{fi}}\tau_n)\Phi(\tau_n - c_v^{\mathrm{fi}}) - \Phi(-\tau_n - c_v^{\mathrm{fi}}) \tag{2.12}$$

Asymptotic Setup: c_v^{as} is the unique solution to

$$\delta = \tau\big[\varphi(c_v^{\mathrm{as}}) - c_v^{\mathrm{as}}\Phi(-c_v^{\mathrm{as}})\big] \tag{2.13}$$

For a derivation of this result see Kohl (2005, Section 10.3).

3. Computation of the finite-sample risk for M-estimators

In this core section of the paper, we present algorithms for the computation of the (maximal) finite-sample risks of the preceding sections. Algorithms A and B can be used for the computation of over-/undershooting risk (2.5), and Algorithms C and D for the maxMSE of (2.4). We also present checks for the accuracy of these algorithms.

3.1. Exact expressions

We fix $n \in \mathbb{N}$ and radius ε_n $[\delta_n] \in [0,1)$ $(*=c)$, $[(*=v)]$. Given some clipping bound $c \in (0,\infty)$, we consider M-estimators S to scores of form (2.7) with equal randomization between the smallest and the largest solutions.

In case of risk (2.5), we are interested in the finite-sample minimax estimator $\tilde{S}_*^{\mathrm{fi}}$, the asymptotic minimax estimator $\tilde{S}_*^{\mathrm{as}}$ and the estimator based on the $\mathrm{O}(n^{-1/2})[\mathrm{O}(n^{-1})]$-corrected $(*=c)$ $[(*=v)]$ asymptotic optimal clipping bound. Suppressing the dependency on the radius ε resp. δ, risk (2.5) of our M-estimator S_n for given $\tau_n \in (0,\infty)$ reads,

$$\mathrm{Risk}\,(S_n;*) = \max\left\{\sup_{Q_{-\tau_n}\in\mathcal{U}_*(-\tau_n)} Q^n_{-\tau_n}(S_n>0), \sup_{Q_{\tau_n}\in\mathcal{U}_*(\tau_n)} Q^n_{\tau_n}(S_n<0)\right\} \tag{3.1}$$

For MSE risk (2.4), we obtain

$$\mathrm{maxMSE}(S_n) = \sup n\int S_n^2\,dQ_n, \qquad Q_n \in \tilde{\mathcal{U}}_c(r;n) \tag{3.2}$$

which we evaluate for M-estimators S_n to scores of form (2.7) where c is one of c^{fo}, c^{so}, c^{to}, c^{o}, c^{FYZ} defined as in Appendix A.2.

3.1.1. Preparations

For any M-estimator S that is based on a score function $\chi_\theta(u)$, which is measurable in u and monotone increasing in $\theta \in \mathbb{R}$, from strictly positive to strictly negative values, the following inclusions hold

$$\{S'>\theta\} \subseteq \big\{ \sum_{i=1}^n \chi_\theta(y_i)>0\big\} \subseteq \{S'\geq\theta\}, \quad \{S''>\theta\} \subseteq \big\{ \sum_{i=1}^n \chi_\theta(y_i)\geq 0\big\} \subseteq \{S''\geq\theta\} \tag{3.3}$$

for any given $y_1, \ldots, y_n$; see Huber (1981, pp. 45). By equivariance, we have $\chi_\theta(y) = \chi_0(y-\theta)$. Therefore, we obtain for any such M-estimator S and any $Q_{-\tau_n} \in \mathcal{U}_*(-\tau_n)$, respectively $Q_{\tau_n} \in \mathcal{U}_*(\tau_n)$,

$$\begin{aligned} Q^n_{-\tau_n}(S>0) &= \tfrac{1}{2}\big[Q^n_{-\tau_n}(S'>0) + Q^n_{-\tau_n}(S''>0)\big] && (3.4)\\ &\leq \tfrac{1}{2}\Big[Q^n_{-\tau_n}\Big(\textstyle\sum_{i=1}^n \chi_0(y_i)>0\Big) + Q^n_{-\tau_n}\Big(\textstyle\sum_{i=1}^n \chi_0(y_i)\geq 0\Big)\Big] && (3.5)\\ &\leq \tfrac{1}{2}\big[Q^n_{-\tau_n}(S'\geq 0) + Q^n_{-\tau_n}(S''\geq 0)\big] && (3.6) \end{aligned}$$

and corresponding relations for $Q^n_{\tau_n}(S<0)$.

3.1.2. Q maximizing over-/undershoooting risk

By monotonicity of χ_0, the probability of $\sum_{i=1}^n \chi_0(y_i) > [\geq]\ 0$ under $Q_{-\tau_n} \in \mathcal{U}_*(-\tau_n)$ is maximal if

$$Q_{-\tau_n}(\chi_0(y)=c) = Q_{-\tau_n}(y\geq c) = Q_0(y\geq c+\tau_n) = \max! \tag{3.7}$$

and $Q_{\tau_n}(\sum_{i=1}^n \chi_0(y_i) < [\leq]\ 0)$ is maximal in $\mathcal{U}_*(\tau_n)$ if $Q_0(y\leq -c-\tau_n)$ is maximal. So we are lead to least favorable pairs $Q'_{-\tau_n} \in \mathcal{U}_*(-\tau_n)$ and $Q''_{\tau_n} \in \mathcal{U}_*(\tau_n)$ which in case $(*=c)$ may be written as

$$Q'_{-\tau_n} = (1-\varepsilon_n)\mathcal{N}(-\tau_n,1) + \varepsilon_n H'_{-\tau_n}, \quad Q''_{\tau_n} = (1-\varepsilon_n)\mathcal{N}(\tau_n,1) + \varepsilon_n H''_{\tau_n} \tag{3.8}$$

with $H'_{-\tau_n}$ and H''_{τ_n} concentrated on $[\tau_n+c,\infty)$ and $(-\infty, -\tau_n-c]$, respectively.
In case $(*=v)$, this leads us to $Q'_{-\tau_n}$ and Q''_{τ_n} having cdf (cf. Rieder (1994, pp. 174))

$$Q'_{-\tau_n}(t) = \big(\Phi(t+\tau_n)-\delta_n\big)_+ + \delta_n H'_{-\tau_n}(t), \quad Q''_{\tau_n}(t) = \big[\Phi(t-\tau_n)+\delta_n H''_{\tau_n}(t)\big] \wedge 1 \tag{3.9}$$

with $H'_{-\tau_n}$ and H''_{τ_n} as in case $(*=c)$.

Remark 3.1. We obtain equality in (3.5) and (3.6) if $Q_\theta \in \mathcal{U}_*(\theta)$ $(\theta \in \mathbb{R})$ is absolutely continuous (a.c.) (cf. Huber (1981, Lemma 10.6.1)). Else, we may have inequalities. However, as shown in Kohl (2005, Remark 11.3.2), for any given non-a.c. $H'_{-\tau_n}$ or H''_{τ_n}, we can specify an a.c. distribution which is stochastically larger, respectively smaller than $H'_{-\tau_n}$ or H''_{τ_n} such that equality holds in (3.5) and (3.6) for this new distribution. Also, there is at least equality in (3.6) even if $H'_{-\tau_n}$ or H''_{τ_n} is not a.c. But, (3.4) could be really smaller than (3.5). That is, to make sure that the finite-sample risk is really attained under $Q'_{-\tau_n}$ and Q''_{τ_n}, (at least) 0 has to be a continuity point of the distributions of S' and S'' under $Q'_{-\tau_n}$ and Q''_{τ_n}. □

3.1.3. Terms for risk calculation

As a first step, we state the distribution of χ_0 under $Q'_{-\tau_n}$ and Q''_{τ_n} where in case $(* = c)$ we have

$$Q'_{-\tau_n}(\chi_0(y) = -c) = (1-\varepsilon_n)\Phi(-c+\tau_n) \tag{3.10}$$

$$Q'_{-\tau_n}(-c < \chi_0(y) < t) = (1-\varepsilon_n)\big[\Phi(t+\tau_n) - \Phi(-c+\tau_n)\big] \quad t \in (-c,c) \tag{3.11}$$

$$Q'_{-\tau_n}(\chi_0(y) = c) = (1-\varepsilon_n)\Phi(-c-\tau_n) + \varepsilon_n \tag{3.12}$$

and correspondingly for Q''_{τ_n} (Kohl, 2005, (11.3.24)–(11.3.26)).
In case $(* = v)$ we get,

$$Q'_{-\tau_n}(\chi_0(y) = -c) = \big(\Phi(-c+\tau_n) - \delta_n\big)_+ \tag{3.13}$$

$$Q'_{-\tau_n}(-c < \chi_0(y) < t) = \big(\Phi(t+\tau_n) - \delta_n\big)_+ - \big(\Phi(-c+\tau_n) - \delta_n\big)_+ \qquad t \in (-c,c) \tag{3.14}$$

$$Q'_{-\tau_n}(\chi_0(y) = c) = 1 - \big(\Phi(c+\tau_n) - \delta_n\big)_+ \tag{3.15}$$

and correspondingly for Q''_{τ_n} (Kohl, 2005, (11.3.30)–(11.3.32)).

3.2. Numerical algorithms for over/-undershooting risk

With the preparations of the preceding subsection we may now formulate our two algorithms for the numerical computation of the finite-sample risk (2.5).

3.2.1. Algorithm A

This procedure directly uses the distribution of χ_0 under $Q'_{-\tau_n}$ / Q''_{τ_n}, which can be read off from equations (3.10)–(3.15). The n-fold convolution of this distribution is calculated using Algorithm 3.4 of Ruckdeschel and Kohl (2010) which is based on fast Fourier transform.

This algorithm wrongly assumes absolute continuity of the distribution of χ_0 under $Q'_{-\tau_n}$ / Q''_{τ_n}. More precisely, it first discretizes the distribution to an equally-spaced grid, applies the Discrete Fourier Transformation (DFT) for convolutions and finally smoothes out the resulting discrete convolution distribution using splines. Despite discretization and smoothing, this algorithm has proven to produce very accurate results already for very small sample sizes (cf. ibid.).

Remark 3.2. Since the law of $\tilde{\psi}_*$ $(* = c, v)$ under $Q'_{-\tau_n}$ and Q''_{τ} puts mass at the points $-c_*, c_*$, for even n, we obtain

$$(Q'_{-\tau_n})^n\Big(\sum_{i=1}^n \tilde{\psi}_*(y_i) > 0\Big) < (Q'_{-\tau_n})^n\Big(\sum_{i=1}^n \tilde{\psi}_*(y_i) \geq 0\Big) \tag{3.16}$$

and similarly for $(Q''_{-\tau_n})^n$. Hence the corresponding non-randomized estimate $\frac{1}{2}(S' + S'')$ is not minimax; see also Huber (1968, Section 4). In contrast, if n is odd, there is no mass at zero and equality holds in (3.16). Anyway, for increasing n the difference between LHS and RHS in (3.16) becomes small very fast as the mass at zero decays exponentially in n. This immediately follows from Hall (1992, Theorem 2.3). □

3.2.2. Algorithm B

Algorithm B in contrast to Algorithm A does take into account the point masses, representing the distributions as mixtures between the a.c. distribution of the restriction Z_i of a normal distribution to interval $[-c, c]$ and a binomial random walk $\sum_i W_i$ with step length c. Since all four cases ($\pm\tau_n$, $* = c, v$) may be treated analogously, we only specify the case $(Q'_{-\tau_n})^n$ and $(* = c)$. To lighten the notation, we define $R := Q'_{-\tau_n}$. In view of (3.10)–(3.12) we can rewrite,

$$R^n\Big(\sum_{i=1}^n \chi_0(y_i) > 0\Big) = R^n\Big(\sum_{i=1}^n \big[(1-V_i)Z_i + V_i W_i\big] > 0\Big) \tag{3.17}$$

with the following stochastically independent random variables

$$V_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bin}\,(1, p_{1,n}), \qquad Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{L}\big(\tilde{Z}_i \,|\, \tilde{Z}_i \in [-c, c]\big), \qquad \tilde{Z}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\,(-\tau_n, 1) \tag{3.18}$$

$$W_i := 2c\tilde{W}_i - c, \qquad \tilde{W}_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bin}\,(1, p_{2,n}) \tag{3.19}$$

where

$$p_{1,n} := (1-\varepsilon_n)\big[\Phi(-c+\tau_n) + \Phi(-c-\tau_n)\big] + \varepsilon_n \tag{3.20}$$

$$p_{2,n} := \big[(1-\varepsilon_n)\Phi(-c-\tau_n) + \varepsilon_n\big]/p_{1,n} \tag{3.21}$$

We abbreviate sums of these random variables of length $m \le n$ by a superscript (m) at the random variable. By stochastic independence we get,

$$\begin{aligned} R^n\Big(\sum_{i=1}^n \chi_0(y_i) > 0\Big) &= R^n\big(Z^{(n)} > 0\big) R^n\big(V^{(n)} = 0\big) + R^n\big(\tilde{W}^{(n)} > n/2\big) R^n\big(V^{(n)} = n\big) + \\ &+ \sum_{j=1}^{n-1} \Big[\sum_{k=0}^{j} R^{n-j}\big(Z^{(n-j)} > c(j-2k)\big) R^j\big(\tilde{W}^{(j)} = k\big)\Big] R^n\big(V^{(n)} = j\big) \end{aligned} \tag{3.22}$$

Of course, we can do the same calculations for $>$ replaced by $\ge$. Thus, the finite simple risk by absolute continuity of the law of $Z^{(m)}$ under R^m reads,

$$\begin{aligned} \mathrm{Risk}(S; c) &= R^n\big(Z^{(n)} > 0\big) R^n\big(V^{(n)} = 0\big) + \tfrac{1}{2}\Big[R^n\big(\tilde{W}^{(n)} > n/2\big) + R^n\big(\tilde{W}^{(n)} \ge n/2\big)\Big] \times \\ &\times R^n\big(V^{(n)} = n\big) + \sum_{j=1}^{n-1}\sum_{k=0}^{j} R^{n-j}\Big(Z^{(n-j)} > c(j-2k)\Big) R^j\big(\tilde{W}^{(j)} = k\big) R^n\big(V^{(n)} = j\big) \end{aligned} \tag{3.23}$$

The m-fold convolution of the law of Z_i is calculated using the FFT-Algorithm as in the case of Algorithm A.

Remark 3.3. If n is even, the law of $\tilde{W}^{(n)}$ puts mass at $n/2$. Thus,

$$R^n\big(\tilde{W}^{(n)} > n/2\big) < R^n\big(\tilde{W}^{(n)} \ge n/2\big) \tag{3.24}$$

and again it follows that the non-randomized estimate $\frac{1}{2}(S' + S'')$ is not minimax; for more details see Remark 3.2. □

3.2.3. Checks

Algorithm B allows to compute the finite-sample risk numerically exactly. This can be checked at least for sample size $n = 2$ where analytic calculations yield

$$R^2\big(Z^{(2)} > 0\big) = 1 - R^2\big(Z^{(2)} \le 0\big) = \\ = 1 - \frac{2\int_{-\sqrt{2}\,c}^{0} \varphi(y + \sqrt{2}\,\tau_2)\Phi(y + \sqrt{2}\,c)\,dy - \Phi(\sqrt{2}\,\tau_2) + \Phi\big(\sqrt{2}\,(\tau_2 - c)\big)}{\big[\Phi(c + \tau_2) - \Phi(-c + \tau_2)\big]^2} \tag{3.25}$$

Moreover, if we choose ε, respectively δ such that the corresponding optimal clipping bound $c_* = \tau_2$ $(* = c, v)$ we obtain

$$R^2(Z^{(2)} \le 0) = \Big[1 - 4\Phi(\sqrt{2}\,c_*)\Phi(-\sqrt{2}\,c_*)\Big] \Big/ \big(2\Phi(2c_*) - 1\big)^2 \tag{3.26}$$

We only compute the finite-sample risk of the asymptotic minimax estimator $\tilde{S}_*^{\mathrm{as}}$ since $c_*^{\mathrm{as}} = \tau_2$ leads to a larger radius than $c_*^{\mathrm{fi}} = \tau_2$ for which Algorithm A is less accurate. The absolute deviation of the finite-sample risk obtained with Algorithm A, respectively Algorithm B from the numerically exact finite-sample risk $\mathrm{Risk}^\natural(\tilde{S}_*^{\mathrm{as}}; *)$ is denoted by error_A resp. error_B. The results are contained in Table 1 where 2^q denotes the number of grid points used in Step 2 of the FFT-Algorithm of Ruckdeschel and Kohl (2010).

$(* = c)$	ε	τ_2	c_c^{as}	$\mathrm{Risk}^\natural(\tilde{S}_c^{\mathrm{as}}; c)$	q	error_A	error_B
	0.0480	2.000	2.000	0.052560	10	5.1e−05	2.4e−08
					12	1.3e−05	1.5e−09
					14	3.2e−06	9.5e−11
	0.2357	1.000	1.000	0.294290	10	8.4e−05	3.9e−08
					12	2.1e−05	2.4e−09
					14	5.2e−06	1.5e−10
$(* = v)$	δ	τ_2	c_v^{as}	$\mathrm{Risk}^\natural(\tilde{S}_v^{\mathrm{as}}; v)$	q	error_A	error_B
	0.0240	2.000	2.000	0.027821	10	2.6e−05	2.6e−08
					12	6.6e−06	1.6e−09
					14	1.6e−06	1.0e−10
	0.1178	1.000	1.000	0.206917	10	3.9e−05	5.6e−08
					12	9.7e−06	3.5e−09
					14	2.4e−06	2.2e−10

Table 1: Precision of the computation for risk (2.5) for $n = 2$

Remark 3.4. **(a)** Algorithms A and B yield small, respectively very small errors for sample size $n = 2$ which underlines the precision of the FFT-Algorithm for both types $* = c, v$.

(b) Huber (1968, p. 278) determines the finite-sample risk of the finite-sample minimax estimator $\tilde{S}_*^{\text{fi}}$ in case of $n = 2$, $\tau_2 = 1.0$ and $\varepsilon_2 = 0.0430$, respectively $\delta_2 = 0.0396$ which by (2.10), respectively (2.12) leads to $c_c^{\text{fi}} = 1.0$, respectively $c_v^{\text{fi}} = 1.0$. Our calculations yield $\text{Risk}^{\natural}(\tilde{S}_c^{\text{fi}}; c) = 0.140359$, respectively $\text{Risk}^{\natural}(\tilde{S}_v^{\text{fi}}; v) = 0.142338$ confirming Huber's results. □

As the FFT-Algorithm maintains its high precision with increasing convolution power, we expect the same behavior as for $n = 2$ for increasing sample size. To check this, we consider the finite-sample risk of the asymptotic minimax estimator in case $* = c$. Moreover, we choose one "typical" situation as the results are almost independent of τ and ε. That is, we fix $\tau = \Phi^{-1}(0.995) \approx 2.576$ and $\varepsilon = 0.2$ which by (2.11) leads to $c_c^{\text{as}} = 1.374$ and determine the distance dist_{AB} between the results of Algorithm A and Algorithm B. Table 2 shows the corresponding results where Risk_{B} denotes the finite-sample risk of S_c^{as} computed with Algorithm B. For yet another cross check for the results of our algorithms, we also calculate empirical confidence intervals by Monte-Carlo simulations. That is, we simulate 1e 06 samples of size n, solve the M equation and compute corresponding 95% confidence intervals for the empirical finite-sample risk. As it turns out, the results of Algorithms A and B always lie well within the 95% confidence interval.

n	emp. risk	95% conf. int.	q	Risk_{B}	dist_{AB}
3	0.0899	[0.0892, 0.0904]	10	0.090029	3.4e−05
			12	0.090029	8.6e−06
			14	0.090029	2.1e−06
5	0.0634	[0.0628, 0.0638]	10	0.063232	3.3e−05
			12	0.063232	8.3e−06
			14	0.063232	2.1e−06
10	0.0387	[0.0382, 0.0391]	10	0.038605	2.0e−05
			12	0.038605	4.9e−06
			14	0.038605	1.2e−06

Table 2: A comparison between Algorithm A and Algorithm B in case ($* = c$).

Remark 3.5. Obviously, Algorithm B is more accurate than Algorithm A (cf. Table 1) but Algorithm A is clearly faster. Moreover, the differences between Algorithm A and Algorithm B are small and decrease quickly with increasing sample size n. Hence, in subsequent computations we use $q = 12$ and Algorithm B for $n \le 10$ and A for $n > 10$. □

3.3. Numerical algorithms for MSE-risk

3.3.1. Q_n maximizing MSE Risk

As shown in Ruckdeschel (2010b), for an M-estimator S_n to scores of form (2.7), in order to maximize MSE up to remainders of order $\text{o}(1/n)$, it suffices to take P^{di} concentrated on $[c + \text{O}(\sqrt{\log(n)/n}), \infty)$. In our examples, a smeared Dirac measure in 100 will largely suffice. In particular, on an event $\tilde{\Omega}_n$ such that the contribution of $\tilde{\Omega}_n^c$ to the MSE-integral is $\text{o}(1/n)$, $\psi \equiv c$ $[P^{\text{di}}]$. Let $\hat{Q}_n \in \tilde{\mathcal{U}}_c(r; n)$ be constructed with an arbitrary such distribution P^{di}.

3.3.2. MSE knowing $\sum U_i = k$

By integration by parts, neglecting $o(1/n)$ remainders,

$$\text{maxMSE}(S_n) \quad = \quad \int S_n^2 \, \hat{Q}_n(S_n \in dx) = \int_0^\infty \hat{Q}_n(|S_n| \geq \sqrt{t}\,)\, dt \tag{3.27}$$

The probability that we have k outliers in $\tilde{\mathcal{U}}_c(r;n)$ is just

$$\rho_{k;n} := P(\textstyle\sum_i U_i = k) = P(\text{Bin}(n, r_n) = k)/P(\text{Bin}(n, r_n) < n/2) \tag{3.28}$$

with $r_n = r/\sqrt{n}$ and hence, under the precautions of Remark 3.1

$$\begin{aligned} \hat{Q}_n(S_n \geq u\,) \quad &= \quad \sum_{1 \leq k < n/2} \hat{Q}_n(\sum_{i:U_i=0} \psi^{(c)}(Y_i - u) + kc > 0, \textstyle\sum_i U_i = k) = \\ &= \quad \sum_{1 \leq k < n/2} \rho_{k;n} \, F^n(\sum_{i=1}^{n-k} \psi^{(c)}(y_i - u) > -kc) \end{aligned} \tag{3.29}$$

3.3.3. Algorithm C

Hence for a u-grid u_ν, $\nu = 1, \ldots, N$ and for $k = 0, \ldots, n/2$, we proceed as in Algorithm B: We change τ_n into $-u_\nu$ in formula (3.18) and obtain in analogy to (3.18) and (3.21)

$$p_{1;\nu} := \big[\Phi(-c - u_\nu) + \Phi(-c + u_\nu)\big], \qquad p_{2;\nu} := \Phi(-c - u_\nu)/p_{1;\nu} \tag{3.30}$$

We correspondingly introduce the random variables

$$Z_{i;\nu} \overset{\text{i.i.d.}}{\sim} \mathcal{L}\big(\tilde{Z}_{i;\nu} \,\big|\, \tilde{Z}_{i;\nu} \in [-c, c]\big), \qquad \tilde{Z}_{i;\nu} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\,(u_\nu, 1) \tag{3.31}$$

$$W_{i;\nu} := 2c\tilde{W}_{i;\nu} - c \qquad \tilde{W}_{i;\nu} \overset{\text{i.i.d.}}{\sim} \text{Bin}\,(1, p_{2;\nu}), \qquad V_{i;\nu} \overset{\text{i.i.d.}}{\sim} \text{Bin}\,(1, p_{1;\nu}) \tag{3.32}$$

and again use the superscript (m)-abbreviation for sums of these random variables. Thus, as in (3.22), we get

$$\begin{aligned} F^n(\sum_{i=1}^{n-k} \psi^{(c)}(y_i - u) > -kc) &= \hat{Q}^n\Big(\sum_{i=1}^{n-k} \big[(1 - V_{i;\nu})Z_{i;\nu} + V_{i;\nu}W_{i;\nu}\big] > -kc\Big) = \\ &= R^n\big(Z_\nu^{(n-k)} > -kc\big) R^n\big(V_\nu^{(n-k)} = 0\big) + R^n\big(\tilde{W}_\nu^{(n-k)} > n/2 - kc\big) R^n\big(V_\nu^{(n-k)} = n - k\big) + \\ &\quad + \sum_{l=1}^{n-1-k} \Big[\sum_{h=0}^{l} R^{n-k-l}\Big(Z_\nu^{(n-k-l)} > c(l - k - 2h)\Big) R^l\big(\tilde{W}_\nu^{(l)} = h\big)\Big] R^n\big(V_\nu^{(n-k)} = l\big) \end{aligned} \tag{3.33}$$

Now for $j = 1, \ldots, n$ and $\nu = 1, \ldots, N$ and $s = -(n-1), \ldots, (n-1)$, by means of the FFT-Algorithm, we determine the triple array

$$R_{\nu;s;j} := R^j\big(Z_\nu^{(j)} > sc\big) \tag{3.34}$$

and then evaluate (3.33) respectively (3.29) term by term.

Remark 3.6. (a) This triple loop over ν, s, j for $R_{\nu;s;j}$ and over k, l, h for (3.29) and (3.33) makes this approach prohibitively slow already for n about 30.

(b) As far as we can see, we cannot avoid this triple loop as the U_i are no longer stochastically independent after having thinned out the neighborhoods. This dependence however is rather weak and may be neglected for $n \geq 30$.

3.3.4. Algorithm D

On the other hand, the contribution of the event $\sum U_i > n/2$ is decaying exponentially, and the same goes for the probability that 0 is a mass point of $\psi^{(c)}$ under $\hat{Q}_n^n$. Hence, even for n about 30 we only make a minor error by neglecting these events. Then, we may directly consider the c.d.f. $R(t)$ of $\psi(Y-u)$ under $\hat{Q}_n$ on a u-grid, which is given by

$$R(t) = \mathrm{I}_{\{-c\le t<c\}}(1-r_n)\Phi(t+u) + \mathrm{I}_{\{t>=c\}} \tag{3.35}$$

and correspondingly its convolutional powers.

3.3.5. Calculation of MSE

Both Algorithms C and D, after having determined $\hat{Q}_n(S_n \ge u)$ on the u-grid, proceed by numerical integration of (3.27) for which we use R-function `integrate` (cf. R Development Core Team (2010)).

4. Higher order approximations of densities/cdf's

As noted before, it is very difficult or even impossible to determine the exact finite-sample risk for sample size $n \ge 3$ analytically. Thus, one might think of higher order approximations of the distributions or densities provided by Edgeworth expansions or saddlepoint approximations to compute (at least) an approximation of the exact finite-sample risk.

4.1. Edgeworth expansions

With some $\chi_0(u)$ of form (2.7), we consider

$$\xi_i = \big(\chi_0(y_i) - \mathrm{E}_R\,\chi_0\big)\big/\sqrt{\mathrm{Var}_R\,\chi_0} \tag{4.1}$$

and assume $R = Q'_{-\tau_n}, Q''_{\tau_n}$ of form (3.8)–(3.9) to be a.c. Thus, $\mathrm{E}_R\,|\xi_i|^5 < \infty$ and the resp. Edgeworth expansion may be read off from Ibragimov (1967, Theorem 1) and Field and Ronchetti (1990, p. 16). That is, we obtain

$$\begin{aligned} &R^n\Big(\sum_{i=1}^n \xi_i < \sqrt{n}\,t\Big) + \mathrm{O}(n^{-3/2}) = \\ &\quad = \Phi(t) - \varphi(t)\bigg[\frac{\rho_R}{6\sqrt{n}}(t^2-1) + \frac{1}{n}\bigg(\frac{\kappa_R}{24}(t^3-3t) + \frac{\rho_R^2}{72}(t^5-10t^3+15t)\bigg)\bigg] \end{aligned} \tag{4.2}$$

where $\rho_R = \mathrm{E}_R\,\xi_1^3$ and $\kappa_R = \mathrm{E}_R\,\xi_1^4 - 3$. To determine an approximation of the finite-sample risk (3.1), we have to choose $t = -\sqrt{n}\,\mathrm{E}_R\,\chi_0\big/\sqrt{\mathrm{Var}_R\,\chi_0}$.

Remark 4.1. If k is even, calculating $\mathrm{E}_R\,\chi_0^k$ ($k \in \mathbb{N}$) in case $(* = v)$, we obtain $\mathrm{E}_{Q'_{-\tau_n}}\,\chi_0^k = \int \chi_0^k\,d\mathcal{N}(-\tau_n,1)$ and $\mathrm{E}_{Q''_{\tau_n}}\,\chi_0^k = \int \chi_0^k\,d\mathcal{N}(\tau_n,1)$, respectively. Probably, this also causes the faster convergence of the finite-sample risks in case of total variation neighborhoods at least for the Edgeworth expansions; see also Remark 5.1. □

4.2. Saddlepoint approximations

Saddlepoint approximations are well known to give excellent accuracy in our context down to very small samples, compare Field and Hampel (1982) and Field and Ronchetti (1990). Assumptions A 4.1–A 4.5 of Field and Ronchetti (1990, Theorem 4.3) are fulfilled for $\chi_t(y) = \chi_0(y-t)$ and $R = Q'_{-\tau_n}, Q''_{\tau_n}$ as defined in (3.8)–(3.9), where we assume R to be absolutely continuous (i.e., $dR = \rho\, d\lambda$). Thus, we can read off an asymptotic expansion of the density of the corresponding M estimator which is

$$f_n(t) = \sqrt{\frac{n}{2\pi}}\, c^{-n}(t) \frac{A(t)}{\sigma(t)} \big[1 + \mathrm{O}(n^{-1})\big] \qquad t \in \mathbb{R} \tag{4.3}$$

where, for $e(t,y) = \exp\{\alpha(t)\chi_t(y)\}$, $c^{-1}(t) = \int e(t,y)\,\rho(y)\,dy$ and

$$\sigma^2(t) = \int \chi_t(y)^2 c(t) e(t,y)\,\rho(y)\,dy, \qquad A(t) = \int \big[\tfrac{\partial}{\partial t}\chi_t(y)\big] c(t) e(t,y)\,\rho(y)\,dy, \tag{4.4}$$

and, for fixed t, $\alpha = \alpha(t)$ is the solution $\alpha \in \mathbb{R}$ to

$$0 = \int \chi_t(y) e(t,y)\,\rho(y)\,dy \tag{4.5}$$

Remark 4.2. (a) One key advantage of saddlepoint approximations for fixed summand distribution, is that once the functions $c(t)$, and $A(t)$ and $\sigma(t)$ of (4.4) are determined on a sufficiently dense grid, this grid is invariant in n. This property gets lost in our shrinking neighborhood setting, where the (least favorable) R is changing with n. So in fact little is gained w.r.t. the FFT algorithm.

(b) Of course, saddlepoint approximations might also be used to fill the triple array $R_{\nu;s;j}$ of (3.34). However, for $n \le 30$ we do not loose too much time in using the FFT-Algorithm instead, while for $n > 30$, we are not sure about the performance of the saddlepoint approximations. □

5. Numerical results

5.1. Over/-undershooting risk

We check the accuracy of the introduced higher order approximations via the very accurate results obtained by our Algorithms A and B. Table 3 shows the results for contamination neighborhoods ($* = c$) where we chose $\tau = \Phi^{-1}(0.95), \Phi^{-1}(0.995) \approx 1.645, 2.576$ and $\varepsilon = 0.1, 0.5$.

For radius $\varepsilon = 0.1$ the saddle point approximation performs well down to sample size 5 or even 3, while for radius $\varepsilon = 0.5$ we need about 5–10 observations to obtain a close approximation. For smaller τ the results of the second-order Edgeworth expansion are comparable with the results of the saddlepoint approximation; i.e., for $\varepsilon = 0.1$ already 3 observations are enough for a reasonable approximation, whereas for $\varepsilon = 0.5$ we need about 5–10 observations. For $\tau = 2.576$ and $\varepsilon = 0.1$ the second-order Edgeworth expansion performs a little worse than the saddle point approximation, whereas for $\tau = 2.576$ and $\varepsilon = 0.5$ the results are again comparable. First-order Edgeworth expansion in all cases yields acceptable results down to sample size 10, but needs 20–50 observations to obtain results comparable to the ones of the second-order expansion and saddle point approximation, respectively.

The corresponding results for total variation neighborhoods ($* = v$) are very similar; see also Table 3.

Remark 5.1. **(a)** Apparently, the speed of convergence towards the asymptotic risk is much faster in case of total variation neighborhoods, which confirms Lemmas A.1 and A.2. Moreover, our numerical calculations yield that the same holds for $\tilde{S}_*^{\text{fi}}$. Tables 3 and 7 indicate that this is also true for $\tilde{S}_*^{\text{as[.c]}}$

(b) Higher order approximations of the clipping height could improve the first-order ones, as the first-order optimal estimator is too optimistic for small sample sizes; for more details see Kohl (2005, Section 11.4), On the other hand, as the computation time of our Algorithm A and even more of our Algorithm B strongly increases with increasing sample size one could also think of saddle point approximations, respectively Edgeworth expansions to determine the finite-sample risk of the finite-sample minimax estimator, especially for larger sample sizes ($n > 20$) since the computation time of these approximations is independent of n. In particular, the results for the Edgeworth expansions can be obtained with very little computational effort. □

5.2. MSE

Under R 2.11.0, we simulated $M = 10000$ runs of sample size $n = 5, 10, 30, 100$ in the ideal location model. In a contaminated situation, we used observations stemming from

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

for $U_i \overset{\text{i.i.d.}}{\sim} \text{Bin}(1, r/\sqrt{n})$, $Y_i^{\text{id}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $Y_i^{\text{di}} \overset{\text{i.i.d.}}{\sim} \text{I}_{\{100\}}$ all stochastically independent and for contamination radii $r = 0.1$. As already indicated in Section 3.3.1 for the considered estimators to be introduced in a moment, a contamination point 100 will largely suffice to attain the maximal MSE on $\tilde{\mathcal{U}}_c(r; n)$.
As estimators we considered M-estimators to scores of type (2.7) with clipping heights $c = 0.7$ (according to the `H07`-estimator from Andrews et al. (1972)) and $c^{\text{fo}}(r) = 1.9483$, the first-order optimal clipping height according to (A.3). All empirical MSE's come with an asymtptotic 95%–confidence interval, which is based on the CLT for the variables

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

To get an idea of the speed of the convergence of the MSE to its asymptotic values, we consider the M-estimators 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 first-order, second-order, and third-order expansion in (2.9). The results are tabulated in Tables 4 and 5 on page 16.

5.3. Finite sample distribution of minimax estimators to over-/undershooting risk

We finally use Algorithm A to compute the cumulative distribution function of $\sum_{i=1}^n \tilde{\psi}_c(y_i)$ under $\left(Q''_{\tau_n}\right)^n$ for different values of n and compare the results with the cumulative distribution function of the normal distribution which has minimum Kolmogorov distance. By symmetry it suffices to consider only the cumulative distribution function of $\sum_{i=1}^n \tilde{\psi}_c(y_i)$ under $\left(Q''_{\tau_n}\right)^n$. Moreover, we give only the two "extreme" situations $\varepsilon = 0.1$ ($\delta = 0.05$), $\tau = 1.645$ (see Figures 1 and 2) and $\varepsilon = 0.5$ ($\delta = 0.25$), $\tau = 2.576$ (see Figure 3 and 4). For the total variation case, in the first situation already 5 observations seem to be enough to get quite close to a normal distribution whereas in the second

$*=c$	ε	τ	c_c^{as}	n	$\mathrm{Risk}^\natural(\tilde{S}_c^{\mathrm{as}};c)$	$\mathrm{Risk}_{\mathrm{EW1}}$	$\mathrm{Risk}_{\mathrm{EW2}}$	$\mathrm{Risk}_{\mathrm{SP}}$
		1.645	1.484	3	12.08	11.83	12.05	12.22
				5	10.62	10.43	10.57	10.57
				10	9.36	9.26	9.35	9.34
				100	7.89	7.88	7.89	7.89
	0.1			asymptotic risk: 7.42				
		2.576	1.675	3	4.75	5.48	4.44	5.16
				5	3.14	3.30	3.02	3.14
				10	1.96	2.01	1.94	1.95
				100	1.08	1.08	1.08	1.08
				asymptotic risk: 0.91				
		1.645	0.663	3	37.10	37.19	38.07	39.36
				5	32.74	32.41	33.10	33.37
				10	28.24	27.84	28.23	28.20
				100	21.28	21.25	21.28	21.28
	0.5			asymptotic risk: 18.74				
		2.576	0.919	3	26.97	27.37	28.84	30.53
				5	20.10	19.90	20.75	21.08
				10	13.42	13.14	13.41	13.40
				100	5.73	5.73	5.73	5.73
				asymptotic risk: 3.98				

$*=v$	δ	τ	c_v^{as}	n	$\mathrm{Risk}^\natural(\tilde{S}_v^{\mathrm{as}};v)$	$\mathrm{Risk}_{\mathrm{EW1}}$	$\mathrm{Risk}_{\mathrm{EW2}}$	$\mathrm{Risk}_{\mathrm{SP}}$
		1.645	1.484	3	9.11	9.06	8.92	8.97
				5	8.40	8.31	8.34	8.32
				10	7.89	7.83	7.88	7.87
				100	7.47	7.46	7.47	7.47
	0.05			asymptotic risk: 7.42				
		2.576	1.675	3	2.62	2.82	2.60	2.66
				5	1.75	1.82	1.72	1.73
				10	1.25	1.29	1.24	1.24
				100	0.93	0.94	0.93	0.93
				asymptotic risk: 0.91				
		1.645	0.663	3	23.47	22.73	24.19	24.23
				5	21.84	21.19	22.00	21.90
				10	20.35	19.98	20.36	20.32
				100	18.90	18.86	18.90	18.90
	0.25			asymptotic risk: 18.74				
		2.576	0.919	3	11.88	12.35	12.79	13.29
				5	8.80	8.83	8.79	8.92
				10	6.20	6.20	6.16	6.18
				100	4.17	4.17	4.17	4.17
				asymptotic risk: 3.98				

Table 3: Approximation quality of Edgeworth expansions and Saddlepoint approximations [Risks are given in percent. $\mathrm{Risk}^\natural(\tilde{S}_*^{\mathrm{as}};*)$ denotes risk (3.1) of $\tilde{S}_*^{\mathrm{as}}$, $\mathrm{Risk}_{\mathrm{EW1}}$ and $\mathrm{Risk}_{\mathrm{EW2}}$ the approximations by means of Edgeworth expansions up to first/second order and $\mathrm{Risk}_{\mathrm{SP}}$ by means of saddlepoint approximations.]

Table 4: emp., num., and as. n maxMSE at $r = 0.1$, $c = c^{\text{fo}} = 1.9483$

$n/$ situation		simulation $\bar{S}_m$	[low; up]	numeric Algo C	Algo D	asymptotics n^0	$n^{-1/2}$	n^{-1}
5	id	0.981	[0.954 ;1.009]	1.008	1.007	1.012	1.012	1.007
	cont	1.471	[1.419 ;1.532]	1.501	1.612	1.054	1.292	1.331
10	id	1.001	[0.973 ;1.029]	1.010	1.009	1.012	1.012	1.010
	cont	1.288	[1.248 ;1.328]	1.290	1.296	1.054	1.222	1.242
30	id	1.028	[1.000 ;1.057]	1.011	1.011	1.012	1.012	1.011
	cont	1.192	[1.158 ;1.226]	1.165	1.167	1.054	1.151	1.158
100	id	0.984	[0.956 ;1.011]	–	1.010	1.012	1.012	1.012
	cont	1.081	[1.050 ;1.111]	–	1.111	1.054	1.107	1.109

Table 5: emp., num., and as. n maxMSE at $r = 0.1$, $c = 0.7$

$n/$ situation		simulation $\bar{S}_m$	[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
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

situation we need a sample size of about 10. This indicates that the speed of convergence in case of total variation neighborhoods is not only faster at zero (finite-sample risk) but uniformly over the whole support of the law of $\sum_{i=1}^n \tilde{\psi}_v(y_i)$ under $\left(Q''_{\tau_n}\right)^n$ than under convex contamination.

Remark 5.2. If H''_{τ_n} is absolutely continuous, then also Q''_{τ_n} is absolutely continuous and by Remark 3.1 the distribution of $\tilde{S}_c^{\text{fi}}$ under $\left(Q''_{\tau_n}\right)^n$ and $\sum_{i=1}^n \tilde{\psi}_c(y_i)$ under $\left(Q''_{\tau_n}\right)^n$ coincide. □

To determine the minimum Kolmogorov distance normal distribution, we use a numerical approximation; i.e., we compute the Kolmogorov distance d_κ of the cumulative distribution functions of $\sum_{i=1}^n \tilde{\psi}_c(y_i)$ under $\left(Q''_{\tau_n}\right)^n$ and $\mathcal{N}(\mu, \sigma^2)$ on a grid of 1e05 points and minimize this distance in μ and σ using the R function `optim`, compare R Development Core Team (2010). As we see, in both cases about 10 observations are enough to get already quite close to a normal distribution where the jumps included in the cumulative distribution functions decay exponentially; see also Remark 3.2.

6. Application: comparison of different "optimal procedures"

6.1. MSE

For the finite-sample MSE risk, for the values of $n = 5, 10, 50, 100$ and $r = 0.1, 0.5, 1.0$, we determine c^{fo}, c^{so}, c^{to}, c^{o}, c^{FYZ} and calculate the MSE-risk for the corresponding M-estimators S.
We compare the resulting IC's w.r.t. their clipping-c and the corresponding numerically exact finite-sample MSE. As the absolute value of MSE is of secondary interest when we want to find the optimal procedure, for all but the procedure with $c = c^{\text{o}}$ we determine the relative MSE, that is

$$\text{relMSE}(\psi^{(c)}) = \text{maxMSE}(\psi^{(c)})/\text{maxMSE}(\psi^{(c^{\text{o}})}) \tag{6.1}$$

For $n = \infty$, we evaluate the corresponding first-order asMSE. As a cross-check the clipping heights c^{fo}, c^{so}, c^{to} are also determined for $n = 10^8$. In case of c_{FZY}, for all finite n's the error tolerance used in `optimize` in R was 10^{-4}, while for $n = \infty$ it was 10^{-12}. For c^{o} and $n = 10^8$, an optimization of the (numerically) exact MSE would have been too time-consuming and has been skipped for this reason. Also, for $n = 5$, the radius $r = 1.0$, corresponding to $\varepsilon = 0.447$, is not admitted for an optimization of the risk proposed by Fraiman et al. (2001) and thus no result is available in this case.

6.2. Over-/undershooting risk

To illustrate the relations between finite-sample and asymptotic results, we numerically check the asymptotics against finite-sample results obtained for fixed neighborhoods. To restrict the amount of results, we choose $\tau = \Phi^{-1}(0.95), \Phi^{-1}(0.995)$ such that 2τ corresponds to the width of 90%, 99%-confidence intervals in case of the standard normal distribution. Moreover, we use $q = 12$ in Algorithms A and B.

We determine the finite-sample risk of the finite-sample minimax estimator $\tilde{S}_c^{\text{fi}}$, the asymptotic minimax estimator $\tilde{S}_c^{\text{as}}$ and the estimator $S_c^{\text{as.c}}$ which is based on the $\text{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound and correspondingly for total variation (with a $\text{O}(n^{-1})$-corrected bound).

Table 6: Optimal clipping heights and corresponding relMSE_n

r		$n=5$	$n=10$	$n=50$	$n=100$	$n=\infty$
0.1	c^{fo}	1.948	1.948	1.948	1.948	1.948
	$\mathrm{relMSE}_n(c^{\mathrm{fo}})$	8.679%	4.065%	0.836%	0.448%	–
	c^{so}	1.394	1.484	1.663	1.724	1.948
	$\mathrm{relMSE}_n(c^{\mathrm{so}})$	0.833%	0.207%	0.014%	0.010%	–
	c^{to}	1.309	1.428	1.644	1.713	1.948
	$\mathrm{relMSE}_n(c^{\mathrm{to}})$	0.332%	0.066%	0.004%	0.006%	–
	c^{FYZ}	1.368	1.370	1.668	1.756	1.939
	$\mathrm{relMSE}_n(c^{\mathrm{FYZ}})$	0.658%	0.002%	0.021%	0.031%	–
	c^{o}	1.167	1.358	1.630	1.704	–
	$\mathrm{MSE}_n(c^{\mathrm{o}})$	1.388	1.239	1.129	1.107	–
0.5	c^{fo}	0.862	0.862	0.862	0.862	0.862
	$\mathrm{relMSE}_n(c^{\mathrm{fo}})$	2.930%	2.655%	0.446%	0.218%	–
	c^{so}	0.650	0.690	0.767	0.790	0.862
	$\mathrm{relMSE}_n(c^{\mathrm{so}})$	0.756%	0.615%	0.036%	0.013%	–
	c^{to}	0.547	0.620	0.744	0.777	0.862
	$\mathrm{relMSE}_n(c^{\mathrm{to}})$	0.230%	0.191%	0.008%	0.003%	–
	c^{FYZ}	0.539	0.632	0.749	0.782	0.866
	$\mathrm{relMSE}_n(c^{\mathrm{FYZ}})$	0.200%	0.248%	0.011%	0.008%	–
	c^{o}	0.413	0.531	0.728	0.770	–
	$\mathrm{MSE}_n(c^{\mathrm{o}})$	4.632	3.039	2.008	1.879	–
1.0	c^{fo}	0.436	0.436	0.436	0.436	0.436
	$\mathrm{relMSE}_n(c^{\mathrm{fo}})$	2.716%	3.132%	0.348%	0.149%	–
	c^{so}	0.320	0.340	0.380	0.394	0.436
	$\mathrm{relMSE}_n(c^{\mathrm{so}})$	1.411%	1.610%	0.076%	0.021%	–
	c^{to}	0.255	0.291	0.361	0.382	0.436
	$\mathrm{relMSE}_n(c^{\mathrm{to}})$	0.876%	0.999%	0.027%	0.006%	–
	c^{FYZ}	–	0.281	0.344	0.387	0.440
	$\mathrm{relMSE}_n(c^{\mathrm{FYZ}})$	–	0.892%	0.063%	0.012%	–
	c^{o}	0.001	0.125	0.334	0.366	–
	$\mathrm{MSE}_n(c^{\mathrm{o}})$	12.627	8.445	4.296	3.787	–

A description to this table is located on page 17.

ε	τ	n	c_c^{fi}	$c_c^{\mathrm{as.c}}$	c_c^{as}	$\mathrm{Risk}_{\mathrm{fi}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as.c}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as}}^{\natural}$
0.1	1.645	3	0.903	0.837	1.484	11.327	11.337	12.079
		5	1.021	0.983		10.188	10.192	10.617
		10	1.148	1.130		9.156	9.156	9.357
		100	1.374	1.372		7.872	7.872	7.888
	2.576	3	0.788	0.621	1.675	3.845	3.871	4.748
		5	0.953	0.859		2.610	2.621	3.143
		10	1.142	1.098		1.751	1.753	1.958
		100	1.497	1.493		1.067	1.067	1.078
0.5	1.645	3	0.198	0.113	0.663	36.695	36.699	37.100
		5	0.287	0.237		32.161	32.167	32.741
		10	0.386	0.362		27.866	27.869	28.244
		100	0.570	0.568		21.239	21.239	21.284
	2.576	3	0.232	0.022	0.919	26.461	26.468	26.966
		5	0.347	0.224		19.006	19.033	20.099
		10	0.487	0.428		12.669	12.683	13.419
		100	0.769	0.763		5.667	5.667	5.731

Table 7: Comparison of optimal clipping bounds and corresponding finite-sample risks in case ($* = c$). [Risks are given in percent.]

Convex contamination: Although there are clear differences between the clipping bounds of the finite-sample and the asymptotic minimax estimator, the differences (in absolute values) concerning the corresponding finite-sample risks are only small; see Figure 5 and Table 7. Moreover, the finite-sample risks of the estimator which is based on the $\mathrm{O}(n^{-1/2})$-corrected asymptotic optimal clipping bound are very close to the finite-sample risk of the finite-sample minimax estimator.

Total variation: The differences between the finite-sample risks (in absolute values) are small already for very small sample sizes ($n \leq 5$); see Figure 6 and Table 8. In particular, the finite-sample risks of $S_v^{\mathrm{as.c}}$ is very close to the finite-sample risk of $\tilde{S}_v^{\mathrm{fi}}$, already for sample size $n = 5$.

7. Conclusion

Appendix A. Details to Minimax Estimators

Appendix A.1. Higher order expansion of MSE in Gaussian context

In the Gaussian case we get:

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

$$A_2 = \tfrac{2}{3}\,\rho_1 {v_0}^3 + {v_0}^4\,(3\,\tilde{v}_2 + l_3) + [\,{v_0}^2\,\big((3\,\tilde{v}_2 + 2\,l_3\,)b^2 + 1\big) + 5\,b^2\,]\,r^2 + \big(\tfrac{l_3}{3}\,b^4 + 3\,b^2\big)\,r^4 \tag{A.2}$$

for

δ	τ	n	c_v^{fi}	$c_v^{\mathrm{as.c}}$	c_v^{as}	$\mathrm{Risk}_{\mathrm{fi}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as.c}}^{\natural}$	$\mathrm{Risk}_{\mathrm{as}}^{\natural}$
0.05	1.645	3	1.120	0.903	1.484	8.879	8.972	9.106
		5	1.232	1.135		8.306	8.321	8.399
		10	1.340	1.310		7.863	7.864	7.891
		100	1.467	1.466		7.465	7.465	7.465
	2.576	3	0.984	0.049	1.675	2.214	2.654	2.618
		5	1.160	0.700		1.588	1.761	1.754
		10	1.351	1.188		1.201	1.215	1.247
		100	1.630	1.627		0.934	0.934	0.935
0.25	1.645	3	0.467	0.394	0.663	23.347	23.360	23.471
		5	0.532	0.502		21.762	21.766	21.843
		10	0.591	0.583		20.326	20.326	20.350
		100	0.655	0.655		18.898	18.898	18.898
	2.576	3	0.477	0.043	0.919	11.559	11.658	11.882
		5	0.595	0.393		8.484	8.589	8.796
		10	0.722	0.656		6.099	6.110	6.203
		100	0.893	0.892		4.168	4.168	4.170

Table 8: Comparison of optimal clipping bounds and corresponding finite-sample risks in case ($* = v$). [Risks are given in percent.]

$$A_c = (2\Phi(c) - 1)^{-1}, \quad b = A_c c, \quad v_0^2 = 2b^2(1 - \Phi(c)) + A_c(1 - 2b\varphi(c))$$
$$l_3 = 2c\varphi(c)/(2\Phi(c) - 1), \qquad \tilde{v}_2 = \frac{6\Phi(c) - 4\Phi(c)^2 - 2 - 2c\varphi(c)}{2c^2(1 - \Phi(c)) + 2\Phi(c) - 1 - 2c\varphi(c)}$$
$$\rho_1 = \frac{3A_c^3\,(1 - 2\Phi(c) + 2c\varphi(c))}{v_0^3} + 3v_0^{-1}$$

for Φ and φ the cumulative distribution function (cdf) of $\mathcal{N}(0,1)$ and its density, respectively.

Appendix A.2. Determining the optimal clipping height

For first-order asymptotics, c^{fo} is determined such that

$$r^2 c^{\mathrm{fo}} = 2\,\mathrm{E}(u - c^{\mathrm{fo}})_+ = 2(\varphi(c^{\mathrm{fo}})) - c^{\mathrm{fo}}\Phi(-c^{\mathrm{fo}})) \tag{A.3}$$

For second-order asymptotics, c^{so} is determined such that

$$r^2 c^{\mathrm{so}} \left(1 + \frac{r^2 + 1}{r^2 + r\sqrt{n}}\right) = 2(\varphi(c^{\mathrm{so}})) - c^{\mathrm{so}}\Phi(-c^{\mathrm{so}})) \tag{A.4}$$

These results may be read off from Rieder (1994, Theorem 5.5.7), and Ruckdeschel (2010a, Corollary 2.2), respectively. For third-order asymptotics, we determine c^{to} by numerical optimization such that the corresponding numerical value of the third-order asymptotic MSE is minimized.

By means of the implicit function theorem, one shows that

$$c^{\mathrm{so}} = c^{\mathrm{fo}} \left(1 - \frac{1}{\sqrt{n}}\,\frac{r^3 + r}{r^2 + 2\Phi(c^{\mathrm{fo}})}\right) + \mathrm{o}(\frac{1}{\sqrt{n}}) \tag{A.5}$$

cf. Ruckdeschel (2010a, equation (2.20)). Additionally, we also determine c^{o} by numerical optimization such that the corresponding numerical value of the (finite-sample) MSE itself is minimized.

Appendix A.3. Approach by Fraiman et al. (2001)

The authors propose risks constructed by means of a function $g : \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}_+$ of asymptotic bias $b(G,\psi)$ and asymptotic variance $v^2(G,\psi)$ where

$$b(G,\psi)/\sqrt{n} := B = \{\beta \,|\, (1-\varepsilon)\textstyle\int \psi_\beta \, dG + \varepsilon b = 0\} \tag{A.6}$$

$$v^2(G,\psi) := \big[(1-\varepsilon)\textstyle\int \psi_B^2 \, dG + \varepsilon b^2\big] \big/ \big[(1-\varepsilon)^2 (\textstyle\int \dot\psi_B \, dG)^2\big] \tag{A.7}$$

The risk of an M-estimator to IC ψ is taken as the function

$$L_g(\psi) = \sup_{G\in\mathcal{V}} g(b(G,\psi), v(G,\psi)/n) \tag{A.8}$$

A mean squared error-type risk is given by $g(u,v) = u^2 + v$. It is not quite the MSE, as it employs the asymptotic terms $b(\cdot)$, $v(\cdot)$. Differently to the scores of type (2.7), the solutions to this problem are of form

$$\psi_{a,b,c,t}(x) = \tilde\psi_{a,b,t}\big(x \min\{1, \tfrac{c}{|x|}\}\big), \quad \tilde\psi_{a,b,t}(x) = a\tanh(tx) + b[x - t\tanh(tx)] \tag{A.9}$$

but the solutions optimal in this approach are numerically quite close (up to 10^{-5}) to corresponding Huber-scores $\psi^{(c)}$.

Appendix A.4. Finitely and asymptotically optimal clipping bound in over-/undershooting risk

Kohl (2005, Lemmas 11.1.3, 11.2.3) derives the following relations between finite-sample-optimal and asymptotically-optimal clipping bounds:

Lemma A.1 ($* = c$). *For $\varepsilon_n \in (0,1)$, $\tau_n \in (0,\infty)$ ($n \in \mathbb{N}$ fixed) it holds,*

$$c_c^{\mathrm{fi}} = c_c^{\mathrm{as}} + \mathrm{O}(n^{-1/2}) \tag{A.10}$$

and the $\mathrm{O}(n^{-1/2})$-corrected optimal asymptotic clipping bound is

$$c_c^{\mathrm{as.c}} := c_c^{\mathrm{as}} - \frac{1}{\sqrt{n}} \frac{\varepsilon(\varepsilon + c_c^{\mathrm{as}}\tau)}{2\tau\Phi(-c_c^{\mathrm{as}})} \tag{A.11}$$

where

$$c_c^{\mathrm{fi}} = c_c^{\mathrm{as.c}} + \mathrm{O}(n^{-1}) \tag{A.12}$$

Lemma A.2 ($* = v$). *For $\delta_n \in (0,1)$, $\tau_n \in (0,\infty)$ ($n \in \mathbb{N}$ fixed) it holds,*

$$c_v^{\mathrm{as}} = c_v^{\mathrm{fi}} + \mathrm{O}(n^{-1}) \tag{A.13}$$

and the $\mathrm{O}(n^{-1})$-corrected asymptotic optimal clipping bound is

$$c_v^{\mathrm{as.c}} = c_v^{\mathrm{as}} - \frac{1}{n} \frac{\tau\big[2\big(c_v^{\mathrm{as}}\big)^2\delta + \tau\varphi(c_v^{\mathrm{as}})\big]}{6\Phi(-c_v^{\mathrm{as}})} \tag{A.14}$$

where

$$c_v^{\mathrm{fi}} = c_v^{\mathrm{as.c}} + \mathrm{O}(n^{-3/2}) \tag{A.15}$$

Remark A.3. Lemma A.1 and Lemma A.2 show that there is a clear difference between contamination and total variation neighborhoods concerning the speed of convergence of the optimal clipping bounds ($\mathrm{O}(n^{-1/2})$ vs. $\mathrm{O}(n^{-1})$). In the mean time, proceeding similarly as Ruckdeschel (2010b), Brandl (2008) has translated the asymptotic expansion for the maximal MSE obtained in the first reference from convex contamination to total variation setting and has obtained an analogue result. As Brandl has found out, this effect is not due to the higher symmetry of the total variation, a first conjecture of the authors of this paper, but rather due to the symmetry of both ψ and $F = \mathcal{N}(0,1)$. The effect also shows up by calculating the Edgeworth expansions; cf. Remark 4.1. □

Acknowledgement

Both authors contributed equally to this work.

References

Andrews D.F., Bickel P.J., Hampel F.R., Huber P.J., Rogers W.H. and Tukey J.W. (1972): *Robust estimates of location. Survey and advances*. Princeton University Press, Princeton, N. J.

Brandl M. (2008): *Higher Order Asymptotics for the MSE of Robust M-Estimators of Location on Shrinking Total Variation Neighborhoods*. Dissertation, Universität Bayreuth, Bayreuth.

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

Field C.A. and Hampel F.R. (1982): Small-sample asymptotic distributions of M-estimators of location. *Biometrika*, **69**: 29–46.

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

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

Huber P.J. (1964): Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**: 73–101.

—— (1968): Robust confidence limits. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **10**: 269–278.

—— (1981): *Robust Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley.

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

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

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*

Rieder H. (1980): Estimates derived from robust tests. *Ann. Statist.*, **8**: 106–115.

—— (1989): A finite-sample minimax regression estimator. *Statistics*, **20**(2): 211–221.

—— (1994): *Robust asymptotic statistics*. Springer Series in Statistics. Springer.

—— (1995): Robustness in structured models. In: Rinne H. (Ed.) *Grundlagen der Statistik und ihre Anwendungen. Festschrift für Kurt Weichselberger.*, p. 172–187. Physica-Verlag, Berlin.

Ruckdeschel P. (2010a): Consequences of Higher Order Asymptotics for the MSE of M-estimators on Neighborhoods. Submitted; ArXiV Nr. 1006.0123.

—— (2010b): Higher Order Asymptotics for the MSE of M-Estimators on Shrinking Neighborhoods. Submitted; ArXiV Nr. 1006.0037.

Ruckdeschel P. and Kohl M. (2010): General Purpose Convolution Algorithm for Distributions in S4-Classes by means of FFT. Submitted; available on the author's web-page.

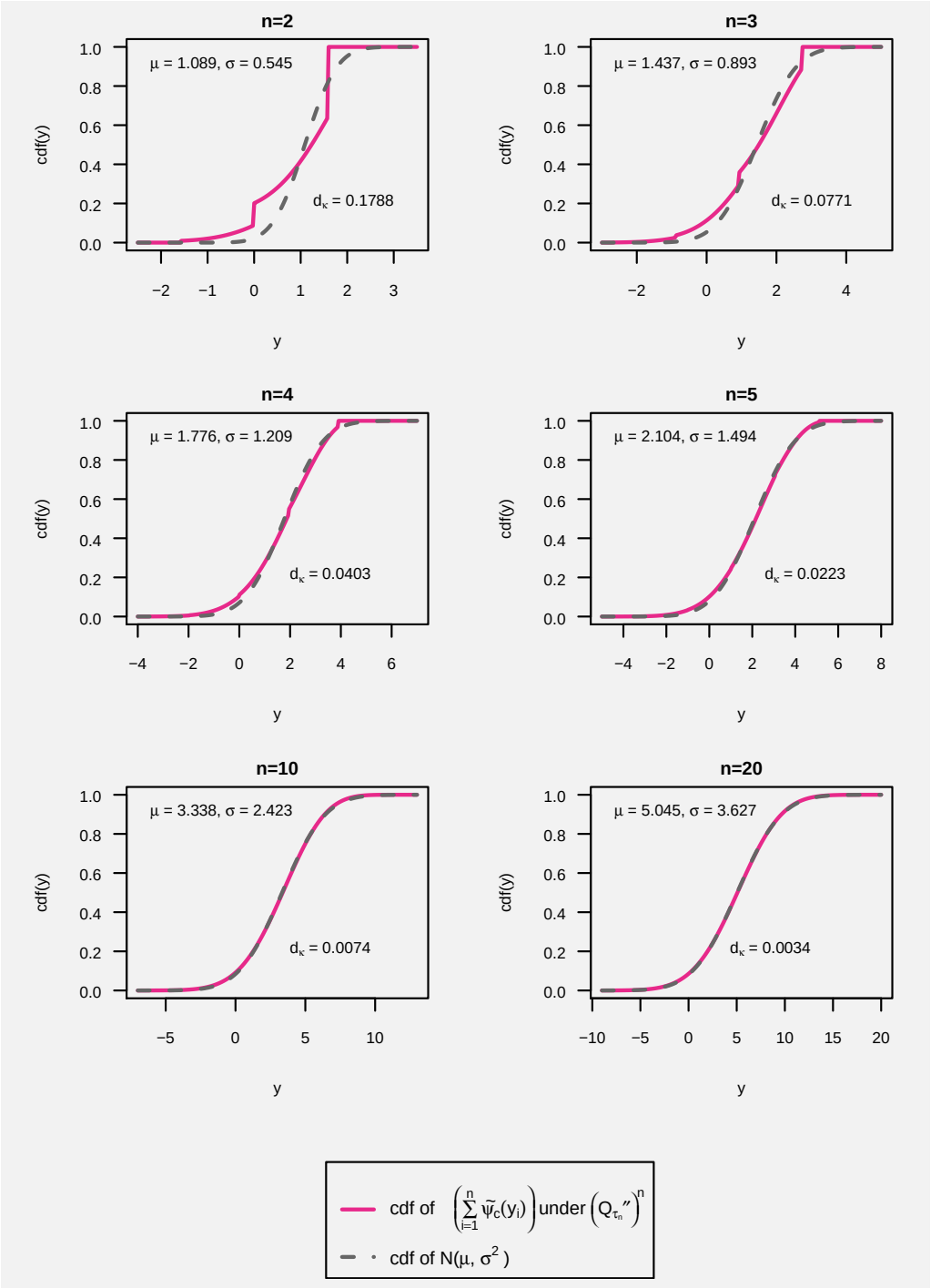

Figure 1: Finite sample distributions for radius $\varepsilon = 0.1$ and $\tau = 1.645$ in case of contamination neighborhoods.

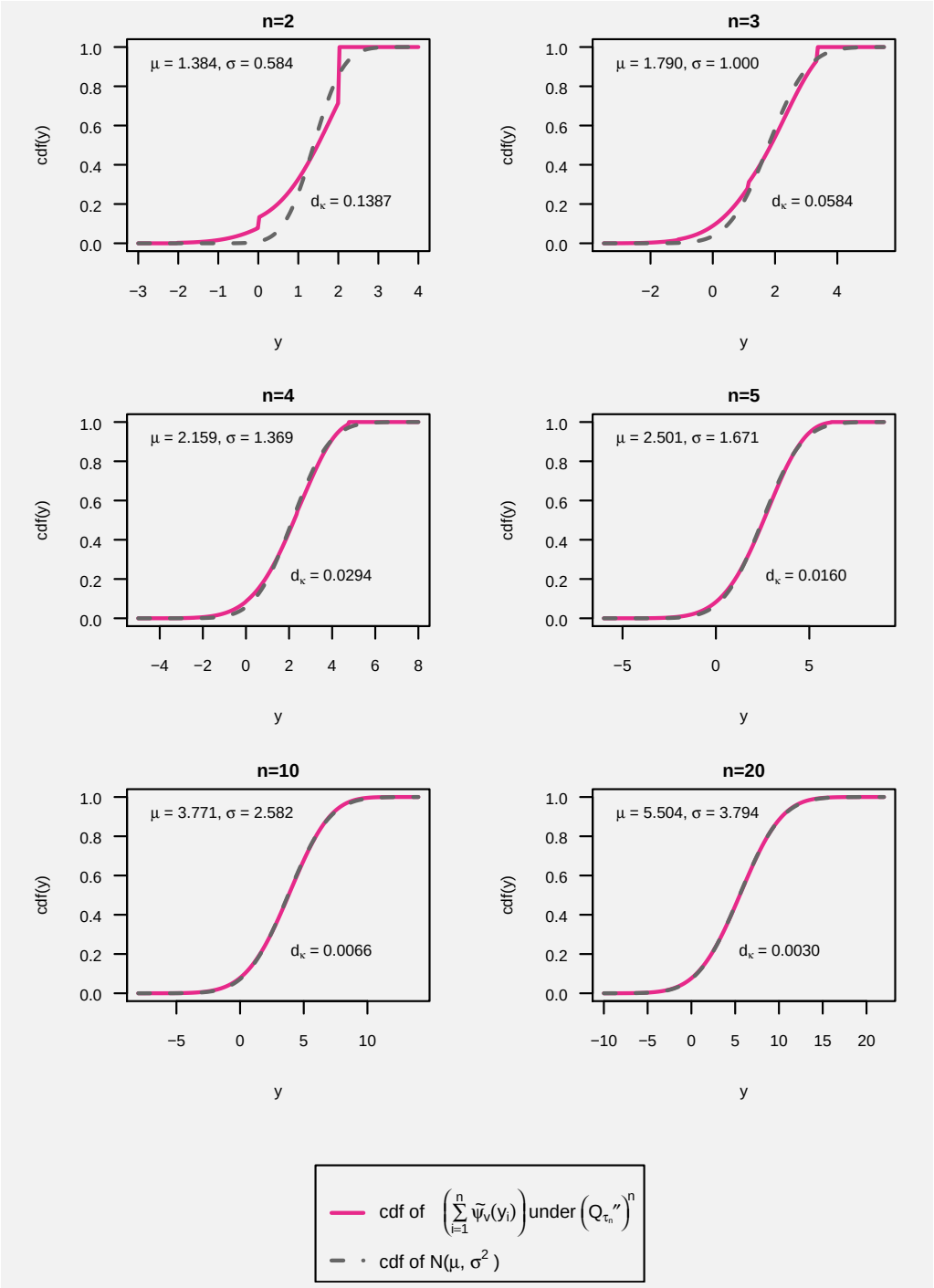

Figure 2: Finite sample distributions for radius $\delta = 0.05$ and $\tau = 1.645$ in case of total variation neighborhoods.

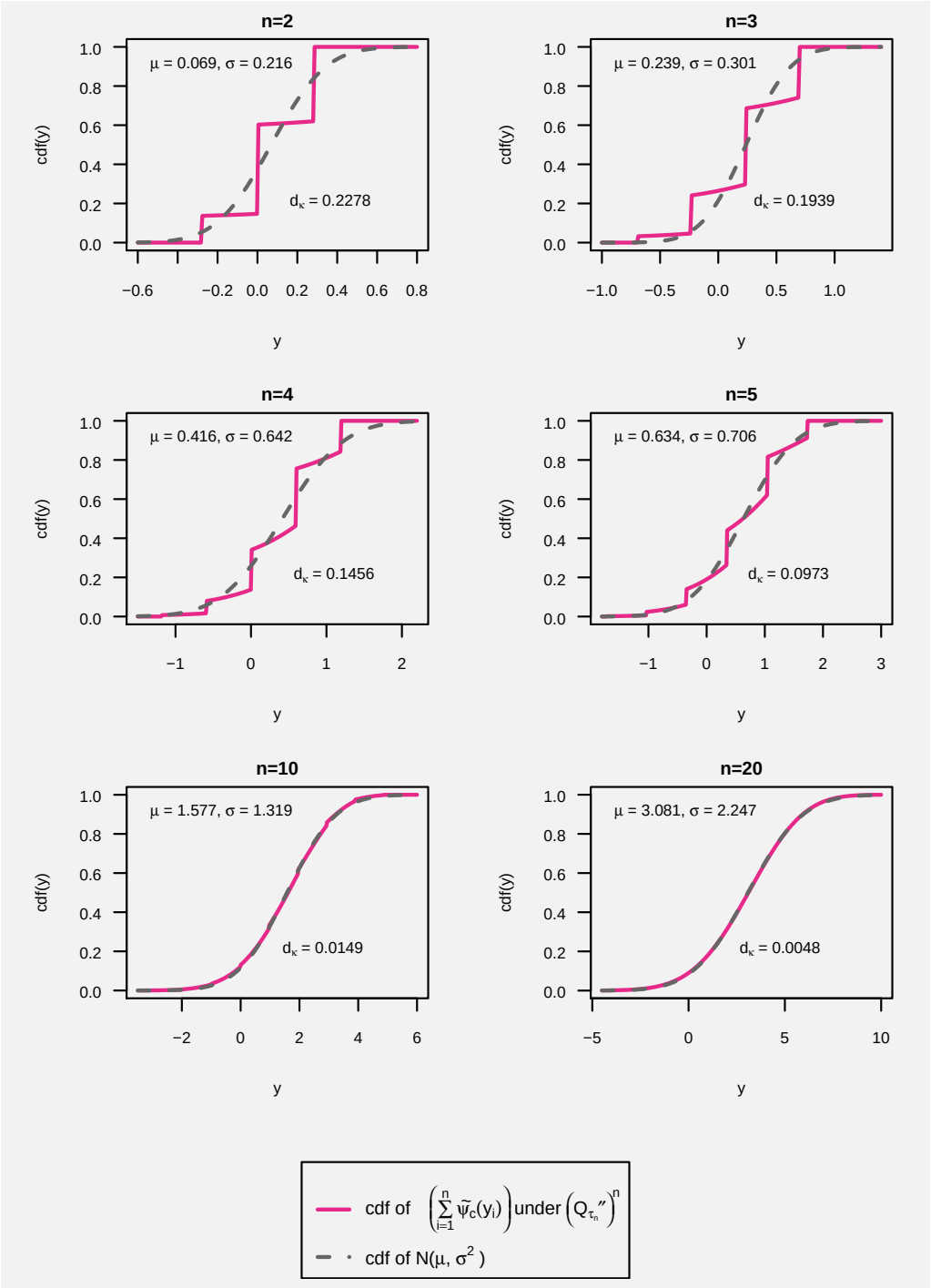

Figure 3: Finite sample distributions for radius $\varepsilon = 0.5$ and $\tau = 2.576$ in case of contamination neighborhoods.

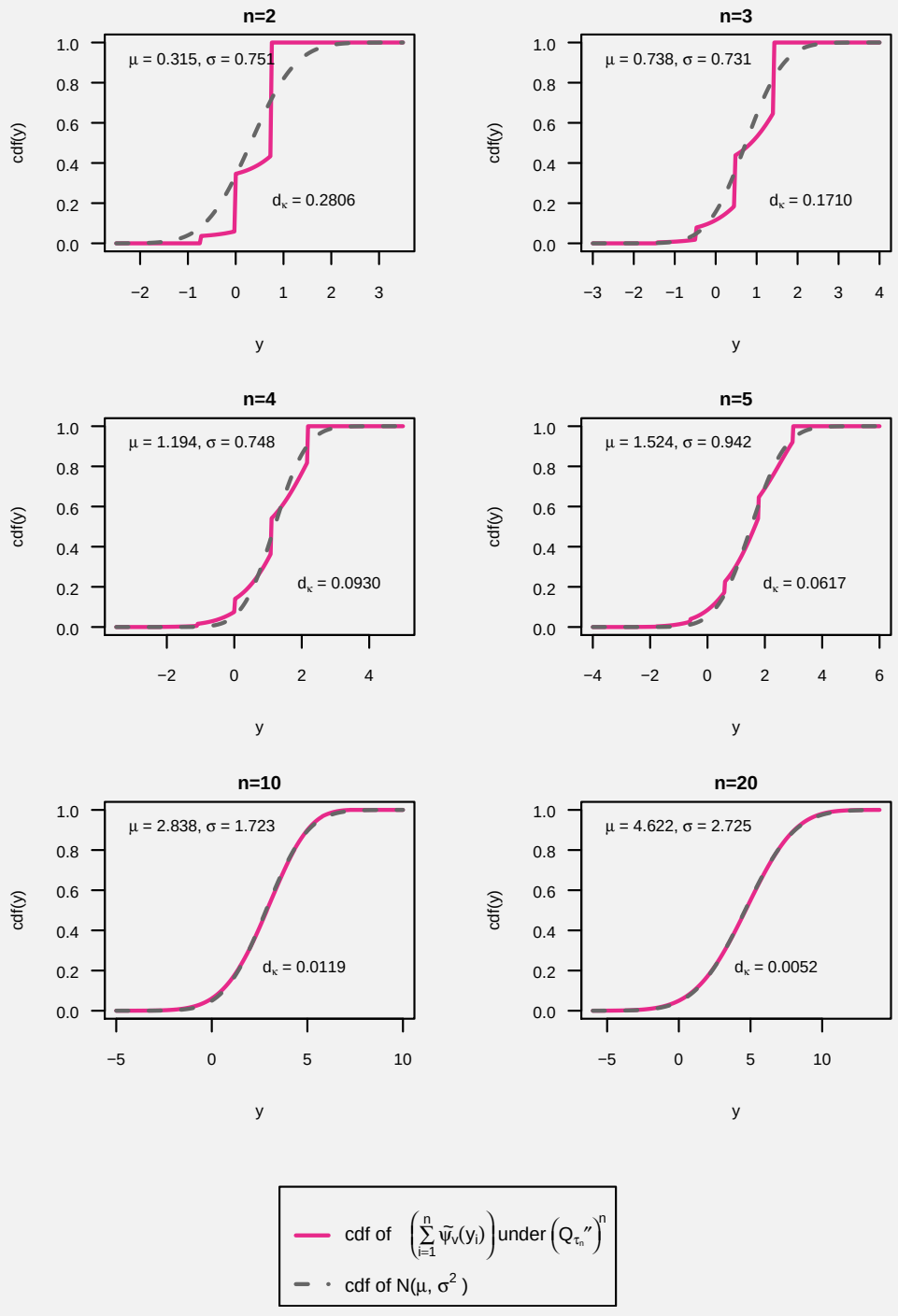

Figure 4: Finite sample distributions for radius $\delta = 0.25$ and $\tau = 2.576$ in case of total variation neighborhoods.

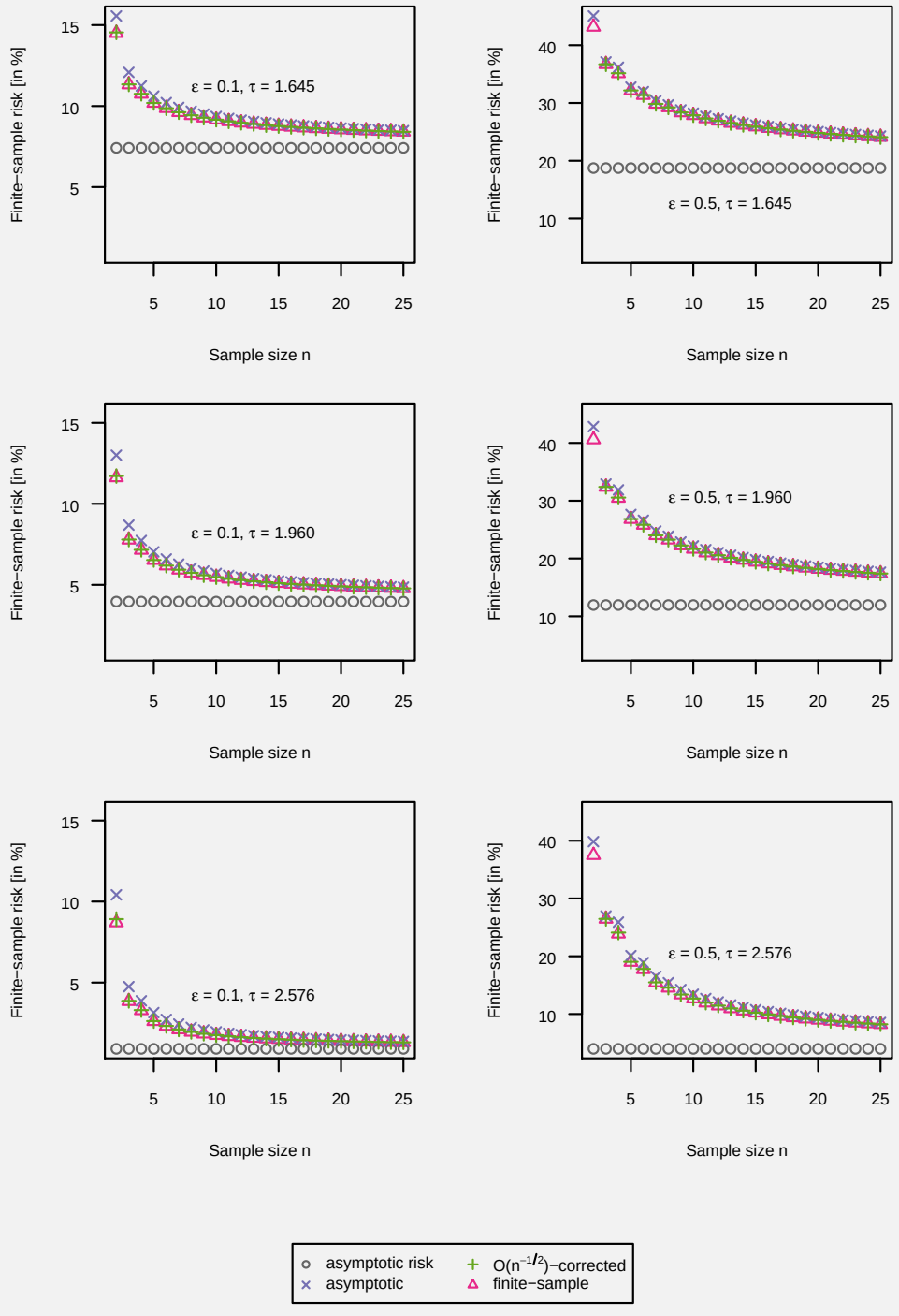

Figure 5: Finite sample risk for sample size $n \leq 25$ given radius $\varepsilon = 0.1, 0.5$ ($* = c$) and width $\tau = 2.576, 1.960, 1.645$.

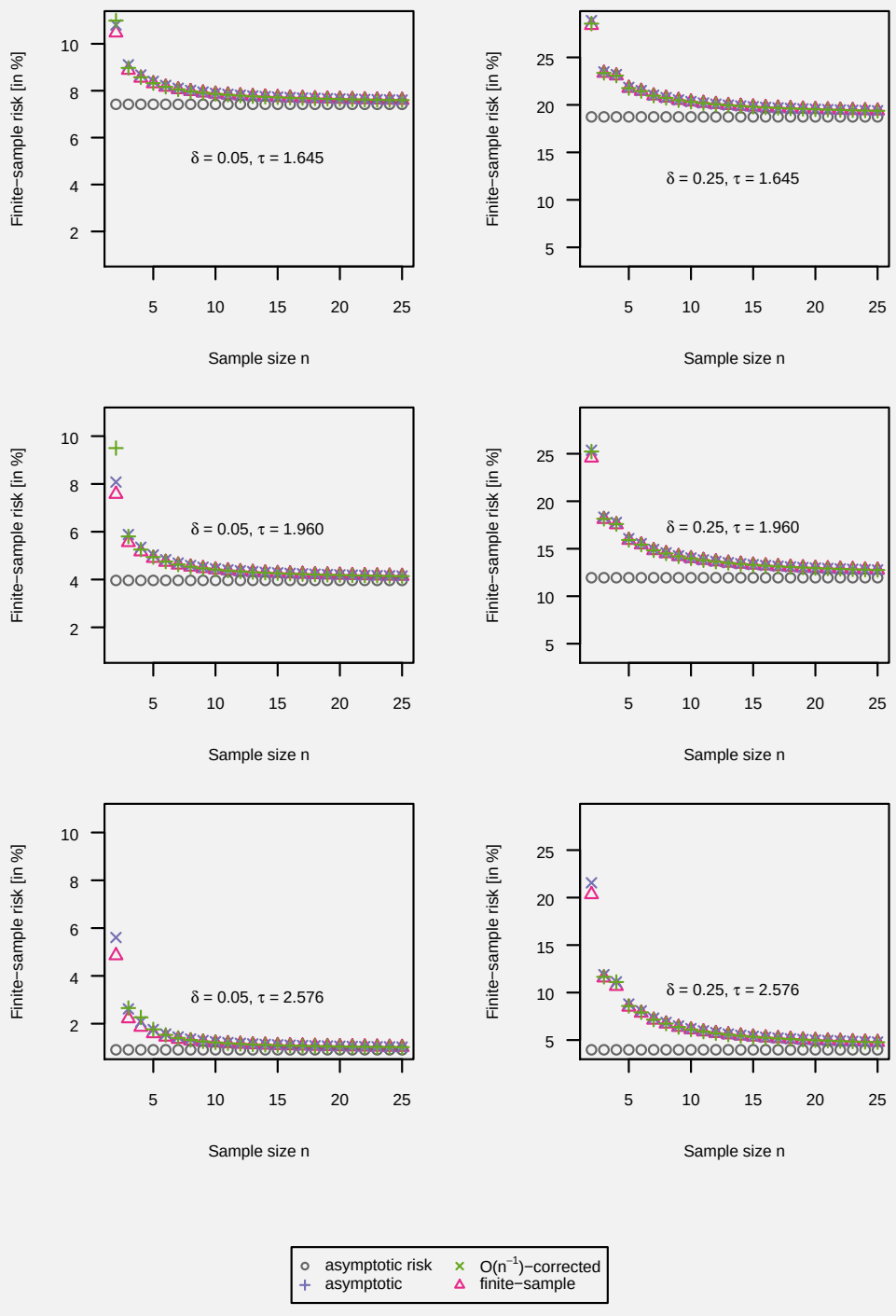

Figure 6: Finite sample risk for sample size $n \leq 25$ given radius $\delta = 0.05, 0.25$ ($* = v$) and width $\tau = 2.576, 1.960, 1.645$.