Higher order asymptotics for the MSE of the sample median on shrinking neighborhoods

Peter Ruckdeschel

Received: date / Accepted: date

Abstract We provide an asymptotic expansion of the maximal mean squared error (MSE) of the sample median to be attained on shrinking gross error neighborhoods about an ideal central distribution. More specifically, this expansion comes in powers of $n^{-1/2}$, for n the sample size, and uses a shrinking rate of $n^{-1/2}$ as well. This refines corresponding results of first order asymptotics to be found in Rieder (1994).
In contrast to usual higher order asymptotics, we do not approximate distribution functions (or densities) in the first place, but rather expand the risk directly.
Our results are illustrated by comparing them to the results of a simulation study and to numerically evaluated exact MSE's in both ideal and contaminated situation.

Keywords sample median · maximal mean squared error · neighborhoods · higher order asymptotics · shrinking neighborhoods · breakdown point

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

1 Motivation/introduction

1.1 Simulations as starting point

This paper was initiated by a simulation study performed by the present author and M. Kohl at Bayreuth university in 2003 for a presentation to be given in the framework of an invitation by S. Morgenthaler to EPF Lausanne. The goal was to investigate the finite sample behavior of procedures, which are distinguished as (first order)

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

asymptotically optimal in infinitesimal robust statistics as to maximal MSE on $\sqrt{n}$-shrinking (convex-contamination) neighborhoods. The results of this study for one dimensional Gaussian location were so promising already for sample sizes n down to about 20 that it seemed worthwhile to dig a little deeper. At closer inspection of the results, we realized that the approximation quality of this first order asymptotics could even be much enhanced down to sample sizes $n = 5$ and 10 if we ignored samples where more than half the sample stemmed from a contamination.

Asymptotically, in our shrinking neighborhood setting, such events carry positive, but exponentially-fast decaying probability for any sample size.

1.2 Description of the main result and discussion

These empirical findings can indeed be substantiated by theory, deriving a uniform higher order asymptotic expansion for the MSE on correspondingly thinned out neighborhoods for the median, location M-estimators for monotone scores, and one-step-constructions.

This paper deals with the median case. It is separated from more general location M-estimators, as the techniques used there are not available for the median due to a failure of a Cramér condition. Although a very particular estimator it reveals that smoothness assumptions on the influence function are not essential to our results, which is of interest more generally.

Moreover, in higher order asymptotics, even for the ideal model, differences appear between diverse variants of the median used for even sample size, a fact which to the author's knowledge has not been spelt out in detail so far.

Denoting by $\tilde{\mathcal{U}}_n(r)$ the neighborhoods thinned out by cutting away samples with more than 50% contaminations, and by M_n a suitable variant of the median, the main result of this paper is

$$\sup_{G^{(n)}\in\tilde{\mathcal{U}}_n(r)} n\,[\mathbf{MSE}(M_n, G^{(n)})] = \frac{1}{4f_0^2}\Big((1+r^2) + \frac{r}{\sqrt{n}}a_1 + \frac{1}{n}a_2\Big) + \mathbf{o}(\tfrac{1}{n}) \qquad (1.1)$$

with a_1 and a_2 certain functions in r, f_0, f_1 and, for a_2, in f_2, where r is the contamination radius and f_i are the values of the ideal density f and its first and second derivatives evaluated at the ideal median.

As a byproduct of the main result, we are able to give necessary and sufficient conditions for a contamination to attain the RHS of (1.1); it is astonishingly small: all mass of the contaminating measures has essentially to be concentrated either left of $-\text{const}\ \sqrt{\log(n)/n}$ or right of $\text{const}\ \sqrt{\log(n)/n}$.

In formula (1.1), we already recognize the following features of the result:
The speed of convergence of the MSE to its asymptotic value is uniform on the whole (modified) neighborhood, and is one order faster in the ideal model; besides, we may work with the original risk (instead of using a modification as usually).

The expansion in powers of $n^{-1/2}$, in the ideal model with first correction term at n^{-1}, comes surprising: Using first order von Mises expansions (compare (1.8) below), in the context of quantiles (comprising the sample median), it can be shown by means of Bahadur-Kiefer representations that the approximation error of this expansion is

an exact $O_{F^n}(n^{-1/4})$—cf. e.g. Jurečková and Sen (1996). So one would expect that under uniform integrability, the first correction term in an expansion of type (1.1) in the ideal model would be of order $n^{-1/4}$, too. In fact, Duttweiler (1973) showed that the L_2-norm of the remainder is of exact order $O(n^{-1/4})$ in our scaled up setup. These results are no contradiction to (1.1), though, as the remainder of course is correlated with the asymptotic linear terms. We still do not see however how Bahadur-Kiefer representations translate into (1.1).
In any case, the approximations of type (1.1) prove very reasonable when compared to both numerical and simulated values of the MSE for finite n.
With the same techniques, we deal with a number of variants of the median for even sample sizes, and specialize these results for the case of $F = \mathcal{N}(0, 1)$. For odd sample size, we also derive asymptotics of this kind for the variance and bias separately.

Remark 1.1 It took some time to write things up in a readable fashion: The proof of the main theorem involves tedious, lengthy asymptotic expansions which are hardly presentable in the framework of an articles—they would slay down any reader by the vast number of terms. Mathematically they are not difficult though and involve basic Taylor expansions at large. Still without the help of a computer algebra software like `MAPLE` for the book-keeping the results would not have been achievable. In the proof section, we hence describe verbally how we got them referring to a corresponding `MAPLE` script available on the web-page to this article. To give you an idea of how tedious terms become, we have included a page of `MAPLE` output on page 24 as a horrifying example.

1.3 Setup

We study the accuracy of the sample median as a location estimator on shrinking neighborhoods: We work in an ideal location model $\mathcal{P} = \{P_\theta \,|\, \theta \in \mathbb{R}\}$ with location parameter θ, observations $X_i \overset{\text{i.i.d.}}{\sim} P_\theta$ and errors u_i given by

$$X_i = u_i + \theta, \qquad u_i \overset{\text{i.i.d.}}{\sim} F \tag{1.2}$$

Due to translation equivariance in the location model we may limit ourselves to $\theta = 0$. We assume that

$$F(0) = 1/2 \tag{1.3}$$

i.e.; the location parameter θ equals the median of the observation distribution, and that F around 0 admits a Lebesgue density f with Taylor expansion about 0 as

$$f(x) = f_0 + f_1 x + \tfrac{1}{2} f_2 x^2 + O(x^{2+\delta_0}), \qquad f_0 > 0 \tag{1.4}$$

for some $\delta_0 > 0$. Furthermore, Finally, we assume that there is a $\delta > 0$ such that

$$\int |x|^\delta f(x)\, dx < \infty \tag{1.5}$$

Remark 1.2 (a) Condition (1.5) is taken from Jurečková and Sen (1982) and is both necessary and sufficient for finiteness of $E_F |M_n|^\gamma$ for any $\gamma > 0$, where M_n is the sample median Med_n to odd sample size n, respectively any variant of the sample median considered in this paper for n even —for a proof see subsection A.1.

(b) By the Hölder-inequality, $\int |x|^\eta F(dx) < \infty$ for each $0 < \eta \le \delta$, so we may assume that $\delta < 1$.

We want to assess both variance and bias simultaneously, so we work with the setup of shrinking neighborhoods as in Rieder (1994), i.e. as deviations from the ideal model (1.2), we consider the set $\mathcal{Q}_n = \mathcal{Q}_n(r)$ of distributions

$$G^{(n)} := \bigotimes_{i=1}^{n} G_{n,i}, \qquad G_{n,i} = (1 - r/\sqrt{n})F + r/\sqrt{n}\, H_{n,i} \tag{1.6}$$

for arbitrary, uncontrollable contaminating distributions $H_{n,i}$. As usual, we interpret $G^{(n)}$ as the distribution of the vector $(X_i)_{i\le n}$ with components

$$X_i := (1 - U_i)X_i^{\mathrm{id}} + U_i X_i^{\mathrm{di}} \tag{1.7}$$

for X_i^{id}, U_i, X_i^{di} stochastically independent, $X_i^{\mathrm{id}} \overset{\text{i.i.d.}}{\sim} F$, $U_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bin}(1, r/\sqrt{n})$, and $(X_i^{\mathrm{di}}) \sim H_n$ for some arbitrary $H_n \in \mathcal{M}_1(\mathbb{B}^n)$. In this setup the median can be understood as an asymptotically linear estimator with influence curve ψ_{Med}, allowing the expansion

$$\mathrm{Med}_n = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} \psi_{\mathrm{Med}}(X_i) + \mathrm{o}_{F^n}(n^{-1/2}), \qquad \psi_{\mathrm{Med}}(x) = \tfrac{\mathrm{sign}(x)}{2f_0} \tag{1.8}$$

— cf. Rieder (1994, Thm. 1.5.1.). Using a clipped version of the quadratic loss function for the estimator $S_n = \mathrm{Med}_n$,

$$\mathrm{MSE}_M(S_n, G) := \mathrm{E}_G(\min(n\,S_n^2, M)), \tag{1.9}$$

we may proceed as outlined in Rieder (1994, p. 207), and obtain

$$\lim_{M\to\infty}\lim_{n\to\infty} \sup_{G^{(n)}}\, n\,[\mathrm{MSE}_M(\mathrm{Med}_n, G^{(n)})] = (4f_0^2)^{-1}(1 + r^2) \tag{1.10}$$

In this paper we want to (a) examine the approximation quality of (1.10), spelling out higher order error terms and (b) discuss the accuracy of this approximation by comparing it to both numerical evaluations of the exact MSE's and simulation results.
In contrast to usual higher order asymptotics, instead of giving approximations to distribution functions (or densities) by Edgeworth expansions or using saddlepoint techniques—cf. e.g. Field and Ronchetti (1990)—we proceed by expanding the risk directly.
As indicated, for (a) we need to modify the neighborhoods, admitting only such samples where less than half of the sample is contaminated, that is $\sum U_i < n/2$ in (1.7). As a side effect of this modification, we will (c) get rid of the somewhat artificial, as statistically unmotivated, modification of the loss function by clipping (1.10), which is common in asymptotic statistics, see, among others, Le Cam (1986), Rieder (1994), Bickel et al. (1998), van der Vaart (1998).

1.4 Organization of the paper

We start with discussing the mentioned modification in detail in section 2. The central theoretical result, Theorem 3.2 is presented in section 3. We then present some ramifications in section 3.2 covering in particular several variants of the sample median for even sample size in Theorem 3.4 and Proposition 3.5; also corresponding

expansions are given for bias and variance separately in Proposition 3.10. Results are spelt out in the special case of $F = \mathcal{N}(0, 1)$ in Corollaries 3.11 and 3.12. These theoretic findings are illustrated with numerical and simulated results in section 4. In the appendix section A, we give proofs to all our assertions.

2 Modification of the shrinking neighborhood setup

The shrinking–neighborhood setup guarantees uniform weak convergence of any as. linear estimator (ALE) to corresponding normal distributions on a representative subclass of the neighboring distributions of form (1.6) — those distributions induced by simple perturbations $Q_n(\zeta, t)$, see Rieder (1994, p. 126).
By the continuous mapping theorem, uniform weak convergence of these ALE's on $\mathcal{Q}_n$ entails uniform convergence of the risk for continuous, bounded loss functions like the clipped version of the MSE (1.9). However, even this (uniform) weak convergence does not entail convergence of the risk for an unbounded loss function like the (unmodified) MSE in general, as we show in the following proposition:

2.1 Convergence failure of the MSE for the median

Proposition 2.1 *Let $\mathcal{P}$ be the location model from* (1.2) *with $f(0) > 0$ and let* Med_n *be the sample median. Then for each odd $n = 2m + 1$ and to any $C > 0$ there is an $x_0 \in \mathbb{R}$ such that with $G_0^{(n)} = [(1 - \frac{r}{\sqrt{n}})F + \frac{r}{\sqrt{n}}\,\mathrm{I}_{\{x_0\}}]^n$*

$$\mathrm{MSE}(\mathrm{Med}_n, G_0^{(n)}) > C \tag{2.1}$$

although, uniformly in $\mathcal{Q}_n$,

$$\sqrt{n}\left(\mathrm{Med}_n - \tfrac{1}{n}\textstyle\sum_{i=1}^n \int \psi_{\mathrm{Med}}\, dG_{n,i}\right) \circ G^{(n)} \xrightarrow{\mathrm{w}} \mathcal{N}(0, (2f(0))^{-2}) \tag{2.2}$$

2.2 Modification of the shrinking neighborhood setup

In view of proposition 2.1, a straightforward modification for finite n consists in permitting only such realizations of $U_1, \ldots, U_n$, where $K = \sum U_i < n/2$. More precisely, for $0 < \varepsilon < 1/2$, we consider the neighborhood system $\tilde{\mathcal{Q}}_n(r, \varepsilon)$ of conditional distributions

$$G^{(n)} = \mathcal{L}\{[(1 - U_i)X_i^{\mathrm{id}} + U_i X_i^{\mathrm{di}}]_i \,\Big|\, \limsup_n \frac{1}{n}\sum U_i \le \varepsilon\} \tag{2.3}$$

If we apply the Hoeffding inequality (Hoeffding (1963, Thm. 2)) to $K = \sum_{i=1}^n U_i$ for the switching variables U_i from (1.7), we obtain

$$P(K > m) \le \exp\left(-2n(\varepsilon - \tfrac{r}{\sqrt{n}})^2\right) \tag{2.4}$$

which shows the announced asymptotic exponential negligibility of this modification. Thus all results on convergence in law of the shrinking neighborhood setup are not

affected when passing from $\mathcal{Q}_n(r)$ to $\tilde{\mathcal{Q}}_n(r)$: Let $B_n := \{K < n/2\}$. Then we have for any $\delta > 0$ and any sequence of events A_n

$$P(A_n \,|\, B_n) = P(A_n \cap B_n)/P(B_n) = P(A_n)(1 + \mathrm{O}(e^{-2n\varepsilon^2/(1+\delta)}))$$

2.3 Connection to the breakdown point

Our definition of the neighborhood $\tilde{\mathcal{Q}}_n(r)$ combines the shrinking neighborhood concept, which will eventually dominate, with a sample–wise restriction; for some number $\varepsilon \in (0, 1)$ depending on the estimator S_n, we only allow for samples where strictly less than $\varepsilon(S_n)n$ observations are contaminated. This number $\varepsilon(S_n)$ is actually just the finite sample (ε-contamination) breakdown point of an estimator S_n introduced by Donoho and Huber (1983).
Thus the concept easily generalizes from the location case to other models: Let $\mathcal{P} = \{P_\theta,\ \theta \in \Theta\}$ be a parametric model and X_i^{id} be $\mathbb{R}^k$-valued observations distributed according to the ideal situation P_θ. We are interested in the question whether for some given estimator S_n, we have uniform convergence of the risk $\int \ell(S_n - \theta)\, dQ_n$ for some loss $\ell \geq 0$ on some (thinned out) neighborhood or not. To this end we define $\tilde{\mathcal{Q}}_n(r, \varepsilon)$ analogously to (2.3). Assume that there is some $\bar{\varepsilon} > 0$ such that for each $n \in \mathbb{N}$ and $k \leq \bar{k} := \lceil n\bar{\varepsilon} \rceil - 1$

$$\varepsilon_0(S_n) := \inf \left\{ \varepsilon^*(X_{n-k}, S_n) \;\middle|\; \begin{array}{c} X_{n-k} = (x_1, \ldots, x_{n-k}) \text{ a possible sample} \\ \text{configuration, } k \leq \bar{k} \end{array} \right\} \geq \bar{\varepsilon} > 0 \tag{2.5}$$

where $\varepsilon^*(X, S)$ is the finite sample (ε-contamination) breakdown point of S at sample X. Then, by an analogue argument to that of Proposition 2.1, the following proposition holds:

Proposition 2.2 *Assume that ℓ is unbounded. Then for any $\varepsilon \geq \varepsilon_0(S_n)$ and $r > 0$, the maximal risk of S_n on $\tilde{\mathcal{Q}}_n(r, \varepsilon)$ is unbounded; in particular, uniform convergence of the risks does not hold.*

The other direction of this connection is more involved and is deferred to a subsequent paper. Under slight additional assumptions, for suitably constructed ALEs to bounded influence curves and for continuous, polynomially growing loss functions, uniform convergence of the risk holds on $\tilde{\mathcal{Q}}_n(r, \varepsilon)$ for any $\varepsilon < \bar{\varepsilon}$. Note that this thinning out for continuous loss functions ℓ is not needed if ℓ is bounded.

3 Higher order asmyptotics for the MSE of the sample median

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

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

$$1 - H_{I_n}^{(n)}((-\infty; c_n]^{i_n}) = \mathrm{o}(d_n) \qquad \left[\, 1 - H_{I_n}^{(n)}((c_n, \infty)^{i_n}) = \mathrm{o}(d_n) \,\right] \tag{3.1}$$

3.1 Main theorem

Theorem 3.2 (a) *In the location model* (1.2) *with ideal central distribution F of finite Fisher information of location, we assume conditions* (1.4) *to* (1.5)*. Then for any $\varepsilon < 1/2$, for $G^{(n)}$ varying in $\tilde{\mathcal{Q}}_n(r,\varepsilon)$ of* (2.3) *it holds*

$$\sup_{G^{(n)}} n\,[\mathrm{MSE}(\mathrm{Med}_n, G^{(n)})] = \frac{1}{4f_0^2}(1 + r^2 + \frac{r}{\sqrt{n}}\,a_1 + \frac{1}{n}\,a_2 + \mathrm{o}(1/n)) \tag{3.2}$$

for

$$a_1 = 2(1+r^2) + \frac{r^2+3}{2}\frac{|f_1|}{f_0^2} \tag{3.3}$$

$$a_2 = (-2 + 3r^2 + 3r^4) + \frac{3r^2(3+r^2)}{2}\frac{|f_1|}{f_0^2} - \frac{3+6r^2+r^4}{12}\frac{f_2}{f_0^3} + \\ + \frac{5(3+6r^2+r^4)}{16}\frac{f_1^2}{f_0^4} \tag{3.4}$$

(b) *The maximal contamination is achieved by any sequence of contaminating measures (H_n), such that for $k_1 > 1$ and $k_2 > \sqrt{5/2}$, and for $\kappa_n = \ulcorner k_1 r\sqrt{n}\,\urcorner$, eventually in n, either*

$$(H_n) \text{ is } \kappa_n\text{–concentrated left of } -\tfrac{k_2}{f_0}\sqrt{\log(n)/n} \text{ up to } \mathrm{o}(n^{-1}) \tag{3.5}$$

or

$$(H_n) \text{ is } \kappa_n\text{–concentrated right of } \tfrac{k_2}{f_0}\sqrt{\log(n)/n} \text{ up to } \mathrm{o}(n^{-1}) \tag{3.6}$$

More precisely, if $f_1 < 0$ [$f_1 > 0$], the maximal MSE is achieved up to $\mathrm{O}(n^{-2})$ by contaminations according to (3.5) *[*(3.6)*], and according to either of the two if $f_1 = 0$.*

Remark 3.3 (a) This result of course also covers the ideal model ($r = 0$), and is also relevant for the fixed neighborhood approach: If for fixed n, we formally plug in $r = s\sqrt{n}$ (for s small in comparison to $\sqrt{n}$) this gives a corresponding result for the maximal MSE of the sample median on a neighborhood of fixed size s. ("formal", as we cannot control the remainder for arbitrary $s < 1$.)

(b) If one is only interested in the behavior of n MSE up to order $\mathrm{o}(n^{-1/2})$, one may weaken assumption (1.4) to: For some $\delta > 0$,

$$f(x) = f_0 + f_1 x + \mathrm{O}(x^{1+\delta}), \qquad f_0 > 0 \tag{3.7}$$

(c) Conditions (3.5) and (3.6) imply that it is sufficient to contaminate F^n by measures H_n the one dimensional marginals of which are either concentrated right of $C\sqrt{\log(n)/n}$ or left of $-C\sqrt{\log(n)/n}$ for some constant $C > 0$ in order to obtain a maximal MSE — an astonishingly modest contamination! With respect to (1.8), this is plausible however, as $|\psi_{\mathrm{Med}}|$ attains its maximal value for any $x \neq 0$.
The thinning out of the marginals by means of Defintion 3.1 even tells us that of the n potentially contaminating X_i^{di} only all subsets of cardinality roughly $\sqrt{n}$ need to be "large" at all, the remaining coset (of cardinality order $n(1 + \mathrm{o}(1))$) of contaminations might even stem from the ideal situation!
As shown in Proposition 3.9, conditions (3.5) resp. (3.6) are almost necessary.

(d) The sample median for odd sample size as well as all variants of the median considered in Proposition 3.4 come up with the same leading term $(1 + r^2)/(4f_0^2)$ for n MSE— according to first order asymptotics (1.10) (with modified loss there!).

(e) In all variants of the sample median considered in Theorem 3.2 and Proposition 3.4, the second order correction is positive, so that for any $r > 0$ we eventually underestimate the MSE by first order asymptotics.

3.2 Ramifications

As simulations in section 4.2 were made for even sample size, we present an analogue to Theorem 3.2 for even sample size below. As there are infinitely many sample medians for even sample size, we consider the following variants:

- the order statistics $X_{[m:n]}$
- the order statistics $X_{[(m+1):n]}$
- the randomized estimator $M_n' := UX_{[m:n]} + (1-U)X_{[(m+1):n]}$ with some randomization $U \sim \mathrm{Bin}(1, 1/2)$
- the midpoint–estimator $\bar{M}_n := (X_{[m:n]} + X_{[(m+1):n]})/2$
- the bias corrected estimator $M_n'' := (X_{[m:n]} + \frac{1}{2n\,f_0})$

Proposition 3.4 *Under the assumptions of Theorem 3.2, for even sample size $n = 2m$, for the sample median variants $X_{[m:n]}$, $X_{[(m+1):n]}$, M_n', $\bar{M}_n$, M_n'', here denoted by M_n generically, for any $\varepsilon < 1/2$, for $G^{(n)}$ varying in $\tilde{Q}_n(r,\varepsilon)$ of* (2.3) *it holds*

$$\sup_{G^{(n)}} n\,[\mathrm{MSE}(M_n, G^{(n)})] = \frac{1}{4f_0^2}\Big\{(1+r^2) + \tfrac{r}{\sqrt{n}}\Big(a_{1,0} + a_{1,1}\tfrac{f_1}{f_0^2}\Big) + \\ + \tfrac{1}{n}\Big(a_{2,0} + a_{2,1}\tfrac{f_1}{f_0^2} + a_{2,2}\tfrac{f_2}{f_0^3} + a_{2,3}\tfrac{f_1^2}{f_0^4}\Big)\Big\} + \mathrm{o}(\tfrac{1}{n}) \tag{3.8}$$

for some real numbers $a_{i,j} = a_{i,j}(M_n)$ which are given in detail in Proposition 3.5. In any variant, the maximal contamination is achieved by contaminating measures H_n according to either condition (3.5) *or* (3.6) *where the distinction between these two is made as in the case of odd sample size.*

Proposition 3.5 *[Specification of the terms $\boldsymbol{a_{i,j}}$] Splitting up $a_{2,0}$, $a_{2,1}$, $a_{2,2}$ according to*

$$a_{2,0} = a_{2,0,r} + a_{2,0,c}, \qquad a_{2,1} = a_{2,1,r} + a_{2,1,c}, \qquad a_{2,2} = a_{2,2,r} + a_{2,2,c} \tag{3.9}$$

we get

(a) Identical terms for all variants*:*

$$a_{2,3} = \tfrac{5(r^4+6r^2+3)}{16}, \qquad a_{2,2,c} = -1/4, \qquad a_{2,2,r} = -\tfrac{(r^4+6r^2)}{12} \tag{3.10}$$

(b) Varying terms in the ideal model*:*

$$\begin{aligned} &a_{2,0,c}(M_n'') = -2, \qquad a_{2,0,c}(\bar{M}_n) = -3 \\ &a_{2,0,c}(X_{[m:n]}) = a_{2,0,c}(X_{[(m+1):n]}) = a_{2,0,c}(M_n') = -1 \end{aligned} \tag{3.11}$$

(c) Remaining $a_{i,j}$ for $\bar{M}_n$, M_n', and M_n''*:*

$$\begin{aligned} &a_{1,0}(M_n'') = a_{1,0}(M_n') = a_{1,0}(\bar{M}_n) = 2(1+r^2), \\ &a_{1,1}(M_n'') = a_{1,1}(M_n') = a_{1,1}(\bar{M}_n) = (r^2+3)\,\mathrm{sign}(f_1)/2, \end{aligned} \tag{3.12}$$

$$a_{2,0,r}(M_n') = a_{2,0,r}(\bar{M}_n) = 3r^4 + 3r^2 = a_{2,0,r}(M_n'') - 2r^2\,\mathrm{sign}(f_1) \tag{3.13}$$

$$\begin{aligned} &a_{2,1,c}(M_n') = a_{2,1,c}(\bar{M}_n) = 0, \quad a_{2,1,c}(M_n'') = 1, \\ &a_{2,1,r}(M_n') = a_{2,1,r}(\bar{M}_n) = \tfrac{3\,r^2(3+r^2)\,\mathrm{sign}(f_1)}{2} = a_{2,1,r}(M_n'') - r^2 \end{aligned} \tag{3.14}$$

(d) Remaining $a_{i,j}$ for $X_{[m:n]}$ and $X_{[(m+1):n]}$:

$$a_{2,1,c}(X_{[m:n]}) = 3/2 = -a_{2,1,c}(X_{[m:n]}) \tag{3.15}$$

For $X_{[m:n]}$ and $X_{[(m+1):n]}$, condition (3.5) *[*(3.6)*] applies if* $4f_0^2 > \ [<] \ -(3+r^2)f_1$ *Correspondingly, let*

$$s' = \begin{cases} 1 \text{ for } X_{[m:n]} \\ -1 \text{ for } X_{[(m+1):n]} \end{cases} \tag{3.16}$$

and

$$s = \operatorname{sign}((3+r^2)f_1 + s'4f_0^2) \tag{3.17}$$

Then the remaining $a_{i,j}$ for $X_{[m:n]}$ and $X_{[(m+1):n]}$ are given by

$$\begin{aligned} a_{1,0} &= 2 + 2s's + 2r^2, \quad a_{2,1,r} = 3s's\Big((3+s)r^2 + r^4\Big)/2 \\ a_{2,0,r} &= 3r^4 + (3+4s)r^2 \end{aligned} \tag{3.18}$$

In case $4f_0^2 = s'(3+r^2)f_1$, both condition (3.5) *and* (3.6) *up to $\mathrm{o}(n^{-2})$ lead to the same MSE.*

Remark 3.6 In case of the sample median for odd sample size,

$$\begin{array}{lll} a_{1,0} = 2(1+r^2), & a_{1,1} = \frac{(r^2+3)\operatorname{sign}(f_1)}{2} & a_{2,0,c} = -2, \\ a_{2,0,r} = 3r^2 + 3r^4, & a_{2,1,c} = 0, & a_{2,1,r} = \frac{3r^2(3+r^2)\operatorname{sign}(f_1)}{2}, \\ a_{2,2,c} = -\frac{1}{4}, & a_{2,2,r} = -\frac{6r^2+r^4}{12}, & a_{2,3} = \frac{5(3+6r^2+r^4)}{16} \end{array}$$

Remark 3.7 It is a well-known consequence of the Jensen inequality that convexity of both loss and admitted estimation (or more generally decision) domain entails that randomization cannot improve an averaged estimator, compare e.g. Witting (1985, (1.2.98), p. 52). This is reflected by the fact that in both ideal and contaminated situation, $\bar{M}_n$ up to $\mathrm{o}(1/n^2)$ has a smaller MSE than M'_n— the only difference arising in term $a_{2,0,c}$.

Remark 3.8 In the ideal model, as shown in Cabrera et al. (1994, Theorem 1), one even has the peculiarity that, in our notation

$$\mathrm{MSE}(\bar{M}_{2m}, F) - \mathrm{MSE}(M_{2m+1}, F) = -\frac{1}{16m^3 f_0^2} + \mathrm{o}(m^{-3}) \tag{3.19}$$

that is, evaluating the sample median at one more observation (from $2m$ to $2m+1$) deteriorates MSE! As our expansion already stops at $\mathrm{o}(m^{-2})$, we cannot reproduce (3.19) to the given exactitude by means of our representations (3.2) and (3.8).
After correcting (minor) typing errors in formulae (2.2), (2.5), and (2.6) in the cited reference, we obtain (3.2) and (3.8) from (2.2) again; for details refer to the web-page to this article.

Conditions (3.5) / (3.6) almost characterize the risk-maximizing contaminations:

Proposition 3.9 *Under the assumptions of Theorem 3.2, let δ_0. Assume that, for $K = \sum_{i=1}^n U_i$ and $k > (1-\delta)r\sqrt{n}$,*

$$\Pr\Big(\sum_{i=1}^n U_i \,\mathrm{I}(X_i^{\mathrm{di}} \le \sqrt{\log(n)/n}/(2f_0)) \ge 1 \,\Big|\, K = k\Big) \ge p_0 > 0 \tag{3.20}$$

Then, eventually in n, no such sequence of contaminations $G_{\mathrm{b}}^{(n)} \in \tilde{\mathcal{Q}}(r)$, can attain the maximal MSE in (3.2) *as in condition* (3.6) *(i.e. with positive bias). More precisely,*

$$\sup_{G^{(n)}} n\,[\mathrm{MSE}(M_n, G^{(n)})] - n\,[\mathrm{MSE}(M_n, G_{\mathrm{b}}^{(n)})] \ge \frac{p_0}{2nf_0\sqrt{2\pi}} + \mathrm{o}(1/n) \tag{3.21}$$

A corresponding relation holds for condition (3.5).

With the same techniques we can also specify which parts of the MSE —up to order $1/n^2$— are due to variance and which are due to bias; to this end let M_n be the sample median and the midpoint estimator $\bar{M}_n$ for odd resp. even sample size.

Proposition 3.10 *In the situation of Theorem 3.2, for contaminating measures H_n as spelt out in* (3.5)*,* (3.6)*, leading to $G_0^{(n)}$ in* (2.3)*, it holds*

$$n\,[\mathrm{Var}(M_n, G_0^{(n)})] \;=\; \frac{1}{4f_0^2}\left\{1 \;+\; \tfrac{r}{\sqrt{n}}\Big(2+|f_1|\Big) + \right.$$
$$\left. +\tfrac{1}{n}\Big(3r^2-(5-(-1)^n)/2 + \tfrac{3|f_1|r^2}{f_0^2} - \tfrac{f_2(r^2+1)}{4f_0^3} + \tfrac{f_1^2(8r^2+7)}{8f_0^4}\Big)\right\} + \mathrm{o}(\tfrac{1}{n}) \tag{3.22}$$

$$\sqrt{n}\,\Big|\mathrm{Bias}(M_n, G_0^{(n)})\Big| \;=\; \frac{1}{2f_0}\left\{r \;+\; \tfrac{1}{\sqrt{n}}\Big(r^2 - \tfrac{|f_1|(r^2+1)}{4f_0^2}\Big) + \right.$$
$$\left. +\tfrac{r}{n}\Big(r^2 - \tfrac{|f_1|\,(r^2+1)}{2f_0^2} + \tfrac{f_2(r^2+3)}{24f_0^3} + \tfrac{f_1^2(r^2+3)}{8f_0^4}\Big)\right\} + \mathrm{o}(\tfrac{1}{n}) \tag{3.23}$$

$$n\,[\mathrm{Bias}^2(M_n, G_0^{(n)})] \;=\; \frac{1}{4f_0^2}\left\{r^2 \;+\; \tfrac{r}{\sqrt{n}}\Big(2r^2 + \tfrac{|f_1|(r^2+1)}{2}\Big) + \right.$$
$$\left. +\tfrac{1}{n}\Big(3r^4 + \tfrac{3|f_1|r^2(r^2+1)}{2f_0^2} - \tfrac{f_2 r^2(r^2+3)}{12f_0^3} + \tfrac{f_1^2(5r^4+14r^2+1)}{16f_0^4}\Big)\right\} + \mathrm{o}(\tfrac{1}{n}) \tag{3.24}$$

We next specialize Theorem 3.2 and Proposition 3.4 for the case of $F = \mathcal{N}(0,1)$ for later comparison to numeric and simulated values.

Corollary 3.11 *In the location model about $F = \mathcal{N}(0,1)$,*

$$\sup_{G^{(n)}} n\,[\mathrm{MSE}(M_n, G^{(n)})] = \frac{\pi}{2}\left\{(1+r^2) + \tfrac{r}{\sqrt{n}}a_{1,0} + \tfrac{1}{n}\Big(a_{2,0} + 2\pi\, a_{2,2}\Big)\right\} + \mathrm{o}(\tfrac{1}{n}) \tag{3.25}$$

Corollary 3.12 *In the location model about $F = \mathcal{N}(0,1)$, in the ideal model*

$$n\,\mathrm{MSE}(\mathrm{Med}_n, F) = \frac{\pi}{2}[1 + (\frac{\pi}{2} + a_{2,0,c})/n] + \mathrm{o}(\tfrac{1}{n}), \tag{3.26}$$

As numerical evaluation of (3.26)*, we get in the three cases:*

$$n\,\mathrm{MSE}(\mathrm{Med}_n, F) \doteq \mathrm{o}(\tfrac{1}{n}) + \begin{cases} 1.5708(1-0.4292/n) & \textit{for } \mathrm{Med}_n,\, M_n'' \\ 1.5708(1+0.5708/n) & \textit{for } X_{[m:n]},\, X_{[(m+1):n]},\, M_n' \\ 1.5708(1-1.4292/n) & \textit{for } \bar{M}_n \end{cases} \tag{3.27}$$

This means: We overestimate $\mathrm{MSE}(\mathrm{Med}_n, F)$ by the first order asymptotics for odd sample size n and with estimator M_n'', and to an even higher degree, if we use $\bar{M}_n$. The risk of estimators $X_{[m:n]}$, $X_{[(m+1):n]}$, M_n' however is underestimated.

4 Illustration of the results

To illustrate the approximation, we consider the case of $F = \mathcal{N}(0,1)$ with a number of numerical evaluations and a small simulation study.

n	num. exact $\mathrm{Var}_n^{\mathrm{id}}$	error of asymptotics 1st/2nd order absolute	1st/2nd order relative	3rd order absolute	3rd order relative
		M_n			
5	1.4341	1.366 E − 1	9.527 %	1.790 E − 3	0.125 %
11	1.5088	6.201 E − 2	4.110 %	7.194 E − 4	0.048 %
101	1.5641	6.687 E − 3	0.428 %	1.174 E − 5	0.001 %
		$X_{[n/2:n]}$, $X_{[(n/2+1):n]}$, M_n'			
6	1.7210	−1.502 E − 1	−8.728 %	−7.715 E − 4	−0.044 %
10	1.6610	−9.022 E − 2	−5.431 %	−5.560 E − 4	−0.033 %
100	1.5798	−8.976 E − 3	−0.568 %	−9.445 E − 6	−0.001 %
		M_n''			
6	1.4776	9.320 E − 2	6.307 %	−1.917 E − 2	−1.297 %
10	1.5106	6.019 E − 2	3.984 %	−7.233 E − 3	−0.479 %
100	1.5641	6.665 E − 3	0.426 %	−7.681 E − 5	−0.005 %
		$\bar{M}_n$			
6	1.2884	−2.823 E − 1	−21.913 %	−9.182 E − 2	−7.126 %
10	1.3832	−1.875 E − 1	−13.557 %	−3.697 E − 2	−2.672 %
100	1.5488	−2.200 E − 2	−1.421 %	−4.472 E − 4	−0.029 %

Table 1 Accuracy of the asymptotics in the ideal model

4.1 Numerical Results in the ideal model

In the ideal model, we have evaluated the integrals numerically, using formulas for the densities in the ideal model to be derived later in section A: g_n for the sample median for odd sample size from (A.5) and g_n for the midpoint estimator for even sample size from (A.48). For the numerical calculations, we have used `R 2.11.0`. Note that the limit up to five digits in this case is 1.5708. Further sample sizes are available on the web-page to this article.

4.2 A simulation study

4.2.1 Simulation design

Under `R 2.11.0`, compare R Development Core Team (2010), we simulated $M = 10000$ runs of sample size $n = 5, 10, 30, 100$ in the ideal location model $\mathcal{P} = \mathcal{N}(\theta, 1)$ at $\theta = 0$. In a contaminated situation, we used observations stemming from

$$G_s^{(n)} = \mathcal{L}\{[(1 - U_i)X_i^{\mathrm{id}} + U_i X_i^{\mathrm{cont}}]_i \Big| \sum U_i \le \ulcorner n/2 \urcorner - 1\}$$

for $U_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bin}(1, r/\sqrt{n})$, $X_i^{\mathrm{id}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, $X_i^{\mathrm{cont}} \overset{\text{i.i.d.}}{\sim} \mathrm{I}_{\{100\}}$ all stochastically independent and for contamination radii $r = 0.1, 0.5, 1.0$. Further results for $n = 30, 50$ and/or $r = 0.25, 0.5$ are available on the web-page to this article. With respect to Remark 3.3 (c), the contamination point 100 will largely suffice to attain the maximal MSE on $\tilde{\mathcal{Q}}_n$.

n	r	sim	[low;	up]	num	n^0	$n^{-1/2}$	n^{-1}
$n = 5$	0.00	1.423	[1.384	;1.464]	1.434	1.571	1.571	1.436
	0.10	1.652	[1.602	;1.701]	1.671	1.587	1.728	1.613
	0.50	3.014	[2.917	;3.111]	3.045	1.963	2.842	3.258
	1.00	4.525	[4.394	;4.655]	4.509	3.142	5.952	8.853
$n = 10$	0.00	1.371	[1.333	;1.410]	1.383	1.571	1.571	1.346
	0.10	1.534	[1.491	;1.578]	1.521	1.587	1.687	1.472
	0.50	2.980	[2.882	;3.078]	2.916	1.963	2.584	2.636
	1.00	5.723	[5.568	;5.879]	5.735	3.142	5.129	6.422
$n = 30$	0.00	1.518	[1.476	;1.560]	1.501	1.571	1.571	1.496
	0.10	1.614	[1.569	;1.659]	1.579	1.587	1.644	1.573
	0.50	2.400	[2.331	;2.469]	2.390	1.963	2.322	2.339
	1.00	5.391	[5.245	;5.538]	5.255	3.142	4.289	4.720
$n = 100$	0.00	1.546	[1.503	;1.589]	1.549	1.571	1.571	1.548
	0.10	1.585	[1.541	;1.629]	1.597	1.587	1.618	1.597
	0.50	2.165	[2.106	;2.223]	2.171	1.963	2.160	2.165
	1.00	4.010	[3.911	;4.108]	3.952	3.142	3.770	3.899

Table 2 Asymptotics compared to numerical and simulational evaluations

rel.err	order	$r = 0.00$	$r = 0.10$	$r = 0.25$	$r = 0.50$	$r = 1.00$
1%	1st order asy.	143	320	2449	10016	40127
	2nd order asy.	143	133	85	124	479
	3rd order asy.	17	17	25	48	124
5%	1st order asy.	29	9	92	406	1629
	2nd order asy.	29	25	10	30	101
	3rd order asy.	7	9	11	20	46

Table 3 Minimal n_0 s.t. for $n \geq n_0$ the relative error using first to third order asymptotics for approximating maxMSE(Med_n) on $\tilde{\mathcal{Q}}_n(r, \varepsilon)$ is smaller than 1% resp. 5%

4.2.2 Results

The simulated results for $n\,\text{MSE}(\text{Med}_n, G_s^{(n)})$ come with an asymptotic 95%–confidence interval, which is based on the CLT for the variable

$$\overline{\text{empMSE}}_n = \tfrac{n}{10000} \sum_j [\text{Med}_n(\text{sample}_j)]^2 \tag{4.1}$$

We compare these results to the corresponding numerical "exact" values and to the asymptotical values for approximation order n^0, $n^{-1/2}$, n^{-1} respectively. For even n we take the midpoint–estimator which is the default procedure in R. For the numerical evaluations we use density formulas from section A: $g_{n,k,k}$ for odd sample size from (A.9) and the integrand from (A.50) for even sample size.

For the ideal situation we had simulation results available for all runs to $r \neq 0$, so the actual sample size for $r = 0$ is 40000.

4.3 Discussion

The numerical results of subsection 4.1 show an excellent approximation quality of our formulas (3.2) and (3.8) with specifications (3.9) to (3.14) in the ideal model. In

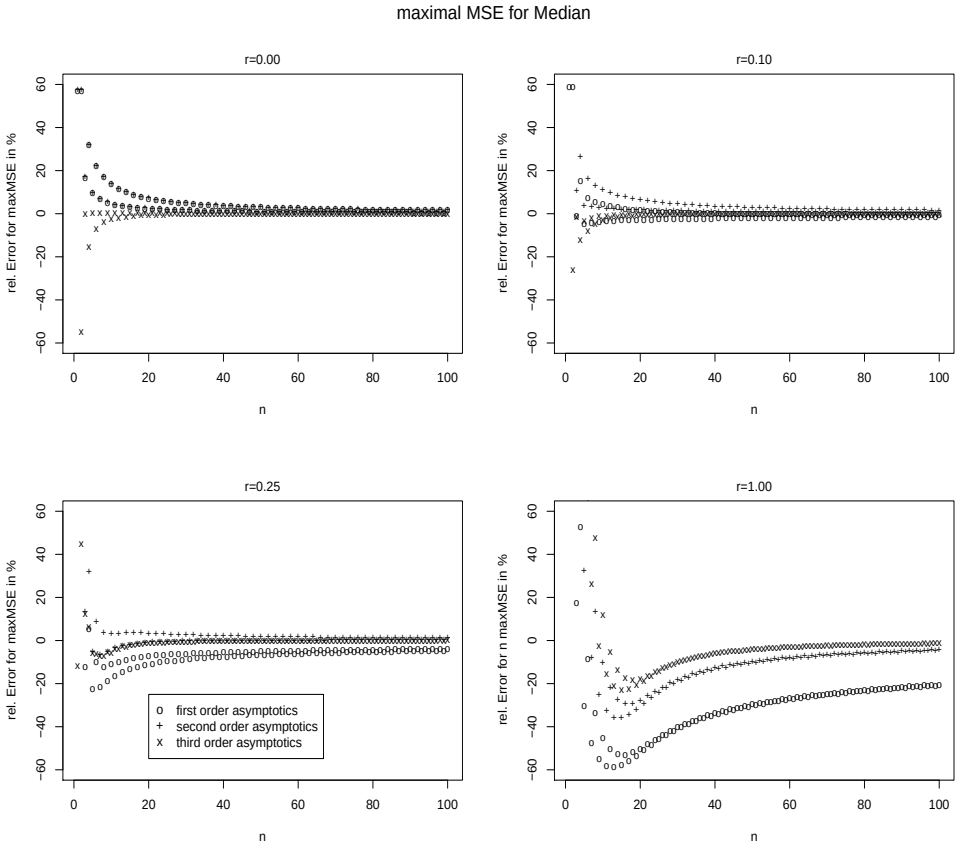

Fig. 1 The mapping $n \mapsto$ rel.error(maxMSE(Med$_n$)) for $F = \mathcal{N}(0, 1)$.

particular the different under/over–estimation properties of the different median variants are closely reflected by the numerical results. The approximation quality of the midpoint estimator indicated in (3.27) is somewhat less well supported by the numerical results, which is probably due to the fact, that by iterated numerical integration the accuracy of the numerical approximation will be inferior to the other variants.
In the contaminated situation, empirical and numerical results also strongly support our assertion of a good approximation quality down to moderate to very small sample sizes, as long as the contamination radius r is not too large: For $n = 5$ upto radius $r = 0.1$, for $n = 10$ (almost) upto $r = 0.25$, for $n > 30$ upto $r = 0.5$, all approximations up to $o(n^{-1})$–terms stay within an (empirical) 95%–confidence interval around the (empirical) MSE (multiplied by n).
In any case, higher order asymptotics yield more accurate approximations than first order ones, and upto case $n = 5$, the $1/n$–terms improve the approximation with respect to the $1/n^{1/2}$–terms.
A closer look is provided by figure 1 (and, zooming in for $n \geq 16$, there is an addi-

tional figure on the web-page). Indeed for all investigated radii $r = 0, 0.10, 0.25, 1.00$, the relative error of our asymptotic formula w.r.t. the corresponding numeric figures is quickly decreasing in absolute value in n; also, we notice a certain oscillation between odd and even sample sizes induced by the different definitions of the sample median in these cases. In table 3, 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 $\mathrm{maxMSE}(\mathrm{Med}_n)$ on $\tilde{\mathcal{Q}}_n(r)$ is smaller than 1% resp. 5% which shows that for $r \leq 0.5$ we need no more than 20 (50) observations to stay within an error corridor of 5% (1%) in third order asymptotics. For first order asymptotics, however we need considerable sample sizes for reasonable approximations unless the radius is rather small.

A Proofs

A.1 Proof of Remark 1.2(a)

Let $n = 2m + 1$ and $\gamma \in (0, 1)$. Necessity: With $\bar{F} = 1 - F$, by integration by parts and Hölder inequality to exponent $m + 1$, we obtain that for any $T > 0$, and some constants $K', K > 0$ and $\alpha = \frac{\gamma-1}{m+1}$

$$\begin{aligned}\mathrm{E}_F\,|\mathrm{Med}_n|^\gamma &= n\binom{2m}{m}\int |x|^\gamma \bar{F}(x)^m F(x)^m\, F(dx) \geq \\ \geq K \max\Big(\int_T^\infty x^{\gamma-1} F(-x)^{m+1}\, dx, \int_T^\infty x^{\gamma-1}\bar{F}(x)^{m+1}\, dx\Big) &\geq K'\Big(\int_{\{|x|>T\}} |x|^\alpha\, F(dx)\Big)^{m+1}\end{aligned}$$

Sufficiency: Under condition (1.5), for $g_\delta(t) = |t|^\delta F(t)\bar{F}(t)$ and $\hat{g}_\delta := \sup_t g_\delta(t)$ it holds —cf. Jurečková and Sen (1982, (2.37))

$$\hat{g}_\delta < \infty, \quad \lim_{|t|\to\infty} g_\delta(t) = 0, \quad I_b := \int [F(t)\bar{F}(t)]^b\, dt < \infty \;\; \forall b \geq 1/\delta \tag{A.1}$$

Hence for any $n > 1 + 2\gamma/\delta$, it follows $b = m + (1 - \gamma)/\delta > 1/\delta$ and hence

$$\begin{aligned}\mathrm{E}_F\,|\mathrm{Med}_n|^\gamma &= n\gamma\binom{2m}{m}\int_0^\infty x^{\gamma-1}[\bar{F}^{m+1}(x)F^m(x) + \bar{F}^m(-x)F^{m+1}(-x)]\, dx \leq \\ \leq n\gamma\binom{2m}{m}\hat{g}_\delta^{(\gamma-1)/\delta}\int_{-\infty}^\infty [\bar{F}(x)F(x)]^b\, dx &\leq n\gamma\binom{2m}{m}\hat{g}_\delta^{(\gamma-1)/\delta} I_b < \infty\end{aligned}$$

The arguments for even sample size are similar. □

A.2 Proof of Proposition 2.1

The assertion for uniform normality follows along the lines of Rieder (1994, Theorem 6.2.8): Although the assumed uniform Lipschitz continuity of the scores ψ—(68), p. 231 in the cited reference—fails, a look into the proof of the theorem shows that this condition only is needed to achieve conclusion $dL(\theta) = \mathbb{I}_k$ on p. 235, which in our situation is the case anyway.
Assertion (2.1) is shown by a breakdown-point argument: We interpret $G^{(n)}$ according to (1.7), where for this proof $X_i \overset{\mathrm{i.i.d.}}{\sim} \mathrm{I}_{\{x_0\}}$. We observe that $\mathrm{Med}_n \geq x_0$ surely under $G^{(n)}$ as soon as $K = \sum U_i$, the number of observations stemming from $\mathrm{I}_{\{x_0\}}$, is larger than m. But, K being a binomial variable, the event $\{K > m\}$ carries positive probability p_n. So setting $x_0 := \sqrt{C/p_n}$, we get
$\mathrm{MSE}(\mathrm{Med}_n, G_0^{(n)}) = \mathrm{E}_{G_0^{(n)}}(\mathrm{Med}_n^2) \geq \mathrm{E}_{G_0^{(n)}}(\mathrm{Med}_n^2\,\mathrm{I}_{\{K>m\}}) \geq x_0^2 p_n = C$ □

A.3 Outline of the proof of Theorem 3.2

As in the theorem we define $n = 2m+1$ and first consider the situation knowing that exactly $K = \sum U_i = k$ observations have been contaminated, to values say $\tilde{x}_1, \ldots, \tilde{x}_k$. More specifically, it will be sufficient to consider—for each fixed t—the number

$$j = j_k(t) := \#\{\tilde{x}_i : \tilde{x}_i \geq t\} \tag{A.2}$$

In this situation we will derive the (conditional) probability that the (unique) median Med_n is not larger than t and derive its density. We then fix some $k_1 > 1$ and $k_2 > \sqrt{5/2}$ and split up the proof according to the following tableau

	$K \leq k_1 r\sqrt{n}$	$k_1 r\sqrt{n} < K \leq \rho n$	$K > \rho n$
$\lvert t\rvert < k_2\sqrt{\log(n)/n}/f_0$	(I)	(III)	excluded
$k_2\sqrt{\log(n)/n}/f_0 \leq \lvert t\rvert \leq n^2$	(II)		
$\lvert t\rvert \geq n^2$	(IV)		

For cases (II) to (IV), we will show that they contribute only terms of order $o(n^{-1})$ to $n\,\mathrm{MSE}(\mathrm{Med}_n)$ 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 $j_k(t)$ either identically k or identically 0 for all t in (I)—or equivalently all $\tilde{x}_i$ are either smaller than $-\frac{k_2}{f_0}\sqrt{\log(n)/n}$ or larger than $\frac{k_2}{f_0}\sqrt{\log(n)/n}$. Integrating out first t and then k we obtain the result.

A.4 $\mathcal{L}(\mathbf{Med}_n)$ in ideal and contaminated situation

A.4.1 Ideal Situation

Lemma A.1 *Let $X_i \overset{\text{i.i.d.}}{\sim} P$ real-valued random variables. Then*

$$P(X_{[k:n]} \leq t) = \sum_{l=k}^{} \binom{n}{l} P(t)^l (1-P(t))^{n-l} \tag{A.3}$$

If $dP = p\,d\lambda$, then $X_{[k:n]}$ has density

$$g(t) = np(t)\binom{n-1}{k-1} P(t)^{k-1}(1-P(t))^{n-k} \tag{A.4}$$

In particular the density of the sample median for odd sample size $n = 2m+1$ in the situation of Theorem 3.2 is

$$g_n(t) = nf(t)\binom{2m}{m} F(t)^m (1-F(t))^m \tag{A.5}$$

Proof The proof is standard, but as we will need some terms later, we pass through the main steps here: For fixed $t \in \mathbb{R}$ we introduce $Y_i := \mathrm{I}_{\{X_i \leq t\}}$. Then the following events are identical

$$\{X_{[k:n]} \leq t\} = \{\#i : \{X_i \leq t\} \geq k\} = \{\sum_{i=1}^{n} Y_i \geq k\} \tag{A.6}$$

The fact that $Y_i \overset{\text{i.i.d.}}{\sim} \mathrm{Bin}(1, P(t))$ entails (A.3). (A.4) follows by simple differentiating, (A.5) by plugging in $k = m+1$. □

A.4.2 Contaminated situation

By (2.3), $X_i = (1 - U_i)X_i^{\mathrm{id}} + U_i X_i^{\mathrm{di}}$, and thus fixing again $t \in \mathbb{R}$, also

$$Y_i = (1 - U_i)Y_i^{\mathrm{id}} + U_i Y_i^{\mathrm{di}} \tag{A.7}$$

with correspondingly defined variables. As we sum up the Y_i in (A.6), only $S_n = \sum Y_i$ will matter. As indicated in the outline, we split up the event $\{S_n > m\}$ by realizations of K, and in the section $\{K = k\}$ we may suggestively write $S_n = S_{n-k}^{\mathrm{id}} + S_k^{\mathrm{di}}$, giving

$$\{\mathrm{Med}_n \le t\} = \dot{\bigcup}_{k=1}^{m} \{S_{n-k}^{\mathrm{id}} + S_k^{\mathrm{di}} > m\} \cap \{K = k\}$$

Splitting up again this event by the realizations of S_k^{di}, we get

$$\{\mathrm{Med}_n \le t\} = \dot{\bigcup}_{k=0}^{m} \dot{\bigcup}_{j=0}^{k} \{S_{n-k}^{\mathrm{id}} > m - j\} \cap \{S_k^{\mathrm{di}} = j\} \cap \{K = k\} \tag{A.8}$$

Thus, for the moment, we may consider the situation that exactly k observations, $0 \le k \le m$, are contaminated, and exactly $j = j_k(t)$ of the contaminated observations are larger than t and denote that event with $D_{j,k,t}$. As $\{X^{id}_{[m-j:n-k]} \le t\}$ is independent from $D_{j,k,t}$, with $\bar{F} = 1 - F$, the the conditional density of Med_n knowing $D_{j,k,t}$ is

$$g_{n,j,k}(t) := (n-k)\binom{2m-k}{m-j} F(t)^{m-j} \bar{F}(t)^{m+j-k} f(t) \tag{A.9}$$

Thus abbreviating again $j_k(t)$ by j, we get the following representation

$$n\,\mathrm{MSE}(\mathrm{Med}_n, G^{(n)}) = n \sum_{k=0}^{m} \sum_{j=0}^{k} \int t^2 g_{n,j,k}(t)\,dt\; P(S_k^{\mathrm{di}} = j)\, P(K = k) \tag{A.10}$$

A.5 Auxiliary results

Before starting with the results we need some preparations

A.5.1 Stirling approximations

We start with writing down some approximations for the factorials and the binomial coefficients derived from the Stirling formula to be found e.g. in Abramowitz and Stegun (1984, 6.1.37):

$$\binom{2n-k}{n-j} = \left(\tfrac{2n-k}{\max(n-j,1)}\right)^{n-j} \left(\tfrac{2n-k}{n+j-k}\right)^{n+j-k} \sqrt{\tfrac{2n-k}{(n+j-k)(n-j)2\pi}}\,(1 + \rho_{n,j,k}), \quad \text{for } -\tfrac{1}{2} - \tfrac{1}{48n} \le \rho_{n,j,k} \le \tfrac{1}{12n} \tag{A.11}$$

$$= \left(\tfrac{2n-k}{n-j}\right)^{n-j} \left(\tfrac{2n-k}{n+j-k}\right)^{n+j-k} \sqrt{\tfrac{2n-k}{(n+j-k)(n-j)2\pi}}\,(1 - \tfrac{1}{8n} + \mathrm{o}(\tfrac{1}{n})), \quad \text{for } j, k = \mathrm{O}(\sqrt{n}), \tag{A.12}$$

The next lemma will be needed to settle case (III):

Lemma A.2 *Let*

$$\kappa := k_1 \log k_1 + 1 - k_1 \tag{A.13}$$

Then it holds that

$$\Pr(\mathrm{Bin}(n, r/\sqrt{n}) > k_1 r \sqrt{n}) \le \exp(-\kappa r \sqrt{n} + \mathrm{o}(\sqrt{n})) \tag{A.14}$$

Proof We first note that $\kappa > 0$, as $\log(x) > 0$ for $x > 1$ and $\kappa = \int_1^{k_1} \log(x)\,dx$. By Hoeffding's inequality (Hoeffding, 1963, Thm. 1, inequality (2.1)), we have for ξ_i, $i = 1, \ldots, n$ i.i.d. real–valued random variables, $|\xi_i| \le M$, $\mu = \mathrm{E}[\xi_1]$ and $0 < \varepsilon < 1 - \mu$

$$P(\tfrac{1}{n}\textstyle\sum_i \xi_i - \mu \ge \varepsilon) \le \left\{ \left(\tfrac{\mu}{\mu+\varepsilon}\right)^{\mu+\varepsilon} \left(\tfrac{1-\mu}{1-\mu-\varepsilon}\right)^{1-\mu-\varepsilon} \right\}^n \tag{A.15}$$

Applying (A.15) to the case of n independent $\mathrm{Bin}(1, r/\sqrt{n})$ variables, we obtain for $B_n \sim \mathrm{Bin}(n, r/\sqrt{n})$ and $0 < \varepsilon = (k_1 - 1)r/\sqrt{n} < 1 - r/\sqrt{n}$:

$$\Pr(B_n > k_1 r\sqrt{n}) \le \exp\Big(- k_1 r\sqrt{n}\,\log(k_1) + (n - k_1 r\sqrt{n})(\log(1 - \frac{r}{\sqrt{n}}) - \log(1 - k_1\frac{r}{\sqrt{n}}))\Big)$$

For $x \in (0,1)$, $-\frac{x}{1-x} \le \log(1-x) \le -x$. Thus the difference of the logarithms is smaller than $(k_1 r)/(\sqrt{n}\,(1 - k_1 r/\sqrt{n})) - r/\sqrt{n}$ and

$$\Pr(B_n > k_1 r\sqrt{n}) \le \exp\big(- \kappa\, r\sqrt{n} + \mathrm{o}(\sqrt{n})\big)$$

□

Corollary A.3 *Let $X \sim \mathrm{Bin}(n, r/\sqrt{n})$. Then for each $i \in \mathbb{N}_0$*

$$\mathrm{E}[X^i\, \mathrm{I}_{\{X \ge k_1 r\sqrt{n}\}}] = \mathrm{o}(n^{-1}) \tag{A.16}$$

Proof $\mathrm{E}[X^i\, \mathrm{I}_{\{X \ge k_1 r\sqrt{n}\}}] \le n^i \Pr(X > k_1 r\sqrt{n}) \overset{(A.14)}{\le} \mathrm{const}\, n^i \exp(-\kappa\, r\sqrt{n})$ □

Lemma A.4 *We have that for $j, k = \mathrm{O}(\sqrt{n})$*

$$|\tfrac{m-j}{2m-k} - F(t)| \le k_2\sqrt{\tfrac{\log(n)}{n}}\,(1 + \mathrm{o}(n^0)) \iff |t| \le \tfrac{k_2}{f_0}\sqrt{\tfrac{\log(n)}{n}}\,(1 + \mathrm{o}(n^0)) \tag{A.17}$$

Proof Using the fact that $j, k = \mathrm{O}(\sqrt{n})$, we note that

$$\tfrac{m-j}{2m-k} = 1/2 + \tfrac{k-2j}{4m} + \tfrac{k(k-2j)}{8m^2} + \mathrm{o}(n^{-1}) \tag{A.18}$$

By (1.4), (1.3), $F(t) = 1/2 + f_0 t + \mathrm{o}(t)$; thus $|\frac{m-j}{2m-k} - F(t)| = |\,\mathrm{O}(\frac{1}{\sqrt{n}}) - f_0 t|$. □

Lemma A.5 *Let $X \sim \mathrm{Bin}(n, p)$. Then, for $p = r/\sqrt{n}$,*

$$\mathrm{E}[X] = rn^{1/2}, \qquad \mathrm{E}[X^2] = r^2 n + rn^{1/2} - r^2, \tag{A.19}$$

$$\mathrm{E}[X^3] = r^3 n^{3/2} + 3r^2 n + (r - 3r^3)n^{1/2} - 3r^2 + 2r^3 n^{-1/2}, \tag{A.20}$$

$$\mathrm{E}[X^4] = r^4 n^2 + 6r^3 n^{3/2} + (7r^2 - 6r^4)n + (r - 18r^3)n^{1/2} + +11r^4 - 7r^2 + 12r^3 n^{-1/2} - 6r^4 n^{-1} \tag{A.21}$$

Proof Cf. the `MAPLE`-procedure `Binmoment` on the web-page. □

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

Lemma A.6 *Let $X \sim \mathcal{N}(0,1)$. Then for $k \in \mathbb{N}$ and any $c > \sqrt{2}$,*

$$\mathrm{E}[|X|^k\, \mathrm{I}_{\{|X| \ge c\sqrt{\log(n)}\}}] = \mathrm{o}(n^{-1}) \tag{A.22}$$

Proof Let $\Phi(x) := \Pr(X \le x)$, $\bar{\Phi} := 1 - \Phi$, $\varphi(x)$ the density of X. Then

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

for some polynomial P_k of degree $k-1$. The assertion follows, as $\varphi(c\sqrt{\log(n)}) = \varphi(0) n^{-c^2/2} = \varphi(0) n^{-(1+\delta)}$ for some $\delta > 0$, and because for the $\Phi(x)$-term, $\bar{\Phi}(x) \le \varphi(x)/x$ for $x > 0$. □

A.6 Proof for odd sample size

We recall the density $g_{n,j,k}$ from (A.9):

$$g_{n,j,k}(t) := (n-k)\binom{2m-k}{m-j} F(t)^{m-j}\bar{F}(t)^{m+j-k} f(t)$$

So the integrand of interest is $n\,t^2\,g_{n,j,k}(t)$. Applying the Stirling approximation (A.12) to the constants, we get

$$\binom{2m-k}{m-j} = \Big(\frac{2m-k}{m-j}\Big)^{m-j}\Big(\frac{2m-k}{m-k+j}\Big)^{m-k+j}\gamma_{n,j,k} \tag{A.23}$$

with

$$\gamma_{n,j,k} := \sqrt{\tfrac{2m-k}{(m+j-k)(m-j)2\pi}}\,(1+\rho_{m,j,k}) \tag{A.24}$$

for $\rho_{m,j,k}$ from (A.12). As $F(t)^{m-j}\bar{F}(t)^{m+j-k}$ suggests an asymptotic decay, we will expand $g_{n,j,k}$ at the mode of $F(t)^{m-j}\bar{F}(t)^{m+j-k}$. Differentiating, we easily get that

$$F(t)^{m-j}\bar{F}(t)^{m+j-k} \le \Big(\frac{m-j}{2m-k}\Big)^{m-j}\Big(\frac{m+j-k}{2m-k}\Big)^{m+j-k} \tag{A.25}$$

with equality iff $t = x_{n,j,k}$ for

$$x_{n,j,k} := F^{-1}(\tfrac{m-j}{2m-k}) \tag{A.26}$$

Introducing

$$\Delta F_{n,j,k} := F(t) - \tfrac{m-j}{2m-k} = F(t) - F(x_{n,j,k}), \tag{A.27}$$

we see that

$$g_{n,j,k}(t) = (n-k)\gamma_{n,j,k} f(t)[1+\tfrac{2m-k}{m-j}\Delta F_{n,j,k}]^{m-j}[1-\tfrac{2m-k}{m+j-k}\Delta F_{n,j,k}]^{m+j-k} \tag{A.28}$$

Case (III): For $k \in [k_1 r\sqrt{n}, \varepsilon n]$, $0 \le j \le k$, we partition the terms according to (A.28) and see that on $|t| \le n^2$, the integrand $n\,t^2\,g_{n,j,k}(t)$ multiplied by n^{-4} is $o(n^0)$ for each fixed t and is dominated by $f(t)$ and hence by dominated convergence tends to 0 as $n \to \infty$. But Lemma A.2 yields that $\Pr(K \ge k_1 r\sqrt{n})$ decays exponentially in n, hence is even $o(n^{-4})$, so as noted, (III) is indeed negligible asymptotically to order $o(n^{-1})$.

Case (II): Here $k \le k_1 r\sqrt{n}$ and $|t| > \frac{k_2}{f_0}\sqrt{\log(n)/n}$, or equivalently by Lemma A.4:

$$|\Delta F_{n,j,k}| > k_2\sqrt{\log(n)/n} \tag{A.29}$$

Now for $x > 0$, $\log(1+x) \le x$ and for $0 < x < 1$, $\log(1-x) \le -x - x^2/2$. Hence, we obtain eventually in n

$$g_{n,j,k}(t)/f(t) \le (n-k)\gamma_{n,j,k}\exp(-\tfrac{(2m-k)^2}{2m}\Delta F_{n,j,k}^2) \overset{\text{(A.12)}}{\underset{\text{(A.29)}}{\le}} (n-k)\sqrt{\tfrac{1}{(m-k/2)}}(1+\tfrac{1}{12m})\exp[-k_2^2\log(m)] \tag{A.30}$$

Plugging in that $m - k/2 \ge m - k_1 r\sqrt{m/2}$ in (II), we get

$$g_{n,j,k}(t)/f(t) \le \text{const}\, m^{\frac{1}{2}-k_2^2}(1+o(n^0)) = o(n^{-2}) \tag{A.31}$$

where the last equality is a consequence of $k_2 > \sqrt{5/2}$. So negligibility follows by dominated convergence.

Case (IV): We only treat the case $t > n^2$; a corresponding relation holds for $t < -n^2$. Under (2.3), for n large enough, we obtain bound $g_{n,j,k} \le n2^n f(t)\bar{F}(t)^{[(1-2\varepsilon)n-1]/2}$. Let $\eta = 1/2 - \varepsilon$, $b = 2/\delta$ and $\delta' \in (0,1)$. By choosing n large enough, we may achieve that $\bar{F}(n^2) =: \lambda_n < 2^{-1/\eta}$ and $F(n^2)^{2b} > 1-\delta'$. So by (A.1), we get eventually in n and for some constant c and any $\eta' > 0$, and g_δ from the proof of Remark 1.2(a)

$$n\int_{n^2}^\infty t^2 g_{n,j,k}(t)\,dt \le \frac{n^2}{1-\delta'}2^n\int_{n^2}^\infty [t^\delta F(t)\bar{F}(t)]^b[F(t)\bar{F}(t)]^b\,dt\lambda_n^{\eta n-1/2-2b} \le$$

$$\le \frac{\hat{g}_\delta^b I_b n^2}{(1-\delta')\lambda_n^{1/2+2b}}(2\lambda_n^\eta)^n = c\exp(-|\log\lambda_n|n[\eta - \tfrac{\log 2}{|\log\lambda_n|} - \tfrac{1+4b}{2n} - \tfrac{2\log n}{n\log\lambda_n}]) \le \exp(-\eta' n)$$

Case (I): Here we restrict ourselves to the case that

$$k \le k_1 \sqrt{n} r, \qquad |\tfrac{m-j}{2m-k} - F(t)| \le k_2 \sqrt{\log(n)/n} \tag{A.32}$$

Doing so, we set $u := t - x_{n,j,k}$. As on (I), $k = \mathrm{O}(\sqrt{n})$ as well as j, we make this magnitude explicit to `MAPLE` in the function `transf` by introducing the bounded variables

$$\tilde{k} := k/\sqrt{m} \qquad \text{and } \tilde{\jmath} := (k/2 - j)/\sqrt{m} \tag{A.33}$$

This gives the expansion in powers of $m^{-1/2}$

$$(m-j)/(2m-k) = 1/2 + \tilde{\jmath}/(4\sqrt{m}) + \tilde{k}\tilde{\jmath}/(8m) + \mathrm{o}(n^{-1}) \tag{A.34}$$

Thus, to get an approximation to $x_{n,j,k} = F^{-1}(\frac{m-j}{2m-k})$, we expand this in a Taylor series in powers of $m^{-1/2}$ (compare our `MAPLE`-procedure `asquantile`) which gives

$$x_{n,j,k} = \tfrac{\tilde{\jmath}}{2f_0\sqrt{m}} + \tfrac{2f_0^2\tilde{\jmath}\tilde{k} - f_1\tilde{\jmath}^2}{8f_0^3 m} + \tfrac{6f_0^4\tilde{\jmath}\tilde{k}^2 - 6f_0^2 f_1\tilde{\jmath}^2\tilde{k} - f_0 f_2\tilde{\jmath}^3 + 3f_1^2\tilde{\jmath}^3}{48 f_0^5 m^{3/2}} + \mathrm{o}(n^{-3/2}) \tag{A.35}$$

Furthermore,

$$f(x_{n,j,k}) = f_0 - \frac{f_1\tilde{\jmath}}{/}(2f_0 m^{1/2}) + (-f_1^2\tilde{\jmath}^2 + 2f_1 f_0^2\tilde{\jmath}\tilde{k} + f_2 f_0\tilde{\jmath}^2)/(8f_0^3 m) + \mathrm{o}(1/n)$$

which implies that in (I), by (A.32), u lies in a shrinking compact, as

$$u = F^{-1}(F(t)) - F^{-1}(F(x_{n,j,k})) = f(x_{n,j,k})^{-1}(F(t) - \tfrac{m-j}{2m-k}) + \mathrm{o}(\sqrt{\log(n)/n}) = \mathrm{O}(\sqrt{\log(n)/n}).$$

Setting $\Delta F_{n,j,k} := F(t) - F(x_{n,j,k})$, and expanding this in a Taylor series around 0, we get

$$\Delta F_{n,j,k} = f_0 u + f_1(u^2/2 + u x_{n,j,k}) + f_2(u^3/6 + (u^2 x_{n,j,k} + u x_{n,j,k}^2)/2) + \mathrm{o}(n^{-3/2})$$

and

$$f(t) = f_0 + f_1(u + x_{n,j,k}) + f_2((u + x_{n,j,k}))^2/2 + \mathrm{o}(n^{-1})$$

We turn to the constant factors now; up to now, the terms arising by applications of the Stirling formulas of subsection A.5.1 come with k–terms in the nominators. As we want to integrate over K later, however, it is preferable to move these terms into the denominators by Taylor approximations —here performed by the functions `asympt` and `collect` in `MAPLE` (compare our function `asbinom`):

$$n(n-k)\sqrt{2\pi}\gamma_{n,j,k} = 2^{\frac{5}{2}} m^{3/2}[1 - \tfrac{\tilde{k}}{4m^{1/2}} + \tfrac{16\tilde{\jmath}^2 - \tilde{k}^2 + 28}{32m}] + \mathrm{o}(n^{\frac{1}{2}}) \tag{A.36}$$

$$\tfrac{(2m-k)^2}{2(m-j)} + \tfrac{(2m-k)^2}{2(m+j-k)} = 4m(1 - \tfrac{\tilde{\jmath}+\tilde{k}}{m^{1/2}} + \tfrac{\tilde{\jmath}^2 + \tilde{\jmath}\tilde{k} + \frac{\tilde{k}^2}{2}}{m}) + \mathrm{o}(n) \tag{A.37}$$

$$\tfrac{(2m-k)^3}{3(m-j)^2} - \tfrac{(2m-k)^3}{3(m+j-k)^2} = \tfrac{16\tilde{k}\sqrt{m}}{3} - (8\tilde{k}^2 + 16\tilde{\jmath}\tilde{k}) + \tfrac{20\tilde{k}^3 + 72\tilde{\jmath}\tilde{k}^2 + 96\tilde{\jmath}^2\tilde{k}}{3m^{1/2}} + \mathrm{o}(n^{-\frac{1}{2}}) \tag{A.38}$$

Next we expand $[1 + \frac{2m-k}{m-j}\Delta F_{n,j,k}]^{m-j}[1 - \frac{2m-k}{m+j-k}\Delta F_{n,j,k}]^{m+j-k}$: We plug in (A.35), set

$$\sigma_n^2 := 8m f_0^2, \qquad y := u\sigma_n \tag{A.39}$$

and apply the Taylor expansion $\exp(x) = 1 + x + x^2/2 + \mathrm{o}(x^2)$. This gives

$$[1 + \tfrac{2m-k}{m-j}\Delta F_{n,j,k}]^{m-j}[1 - \tfrac{2m-k}{m+j-k}\Delta F_{n,j,k}]^{m+j-k} = \exp(-y^2/2)\, h(y, \tilde{\jmath}, \tilde{k}, n) + \mathrm{o}(n^{-1})$$

with

$$h(y, \tilde{\jmath}, \tilde{k}, n) = 1 + \left(\left(\tfrac{\tilde{k}}{4} - \tfrac{f_1\tilde{\jmath}y^2}{2{f_0}^2}\right)y^2 - \tfrac{f_1}{8{f_0}^2}\sqrt{2}y^3\right)m^{-1/2} + P(y, \tilde{k}, \tilde{\jmath})m^{-1} \tag{A.40}$$

where P is some polynomial depending on f_0, f_1, f_2 with $\deg(P)(y) = 6$ the exact expression of which may be drawn from the MAPLE-script. Accordingly, we define $\tilde{x}_{n,j,k} := x_{n,j,k}\sigma_n$, and, with φ the density of $\mathcal{N}(0,1)$, use the abbreviations

$$\tilde{\varphi}(t) = \varphi \circ y \circ u(t), \qquad \tilde{h}(t, \tilde{\jmath}, \tilde{k}, n) = h(y \circ u(t), \tilde{\jmath}, \tilde{k}, n) \tag{A.41}$$

We also introduce the integration domains

$$A_{n,j,k} = \left\{ t \in \mathbb{R} \,\middle|\, |\tfrac{m-j}{2m-k} - F(t)| \le k_2 \sqrt{\tfrac{\log(n)}{n}} \right\}, \qquad \tilde{A}_{n,j,k} = \left\{ |t| \le k_2 \sqrt{\tfrac{\log(n)}{n}} (1 + \mathrm{o}(n^0))/f_0 \right\} \tag{A.42}$$

Finally, applying (A.12) and (A.36), we derive an integration constant $c_{n,j,k}$ from $\gamma_{n,j,k}$ from (A.24):

$$c_{n,j,k} := 2^{-\frac{5}{2}} m^{-\frac{3}{2}} \gamma_{n,j,k} = 1 - \tilde{k}/(4m^{1/2}) + (16\tilde{\jmath}^2 - 16\tilde{\jmath}\tilde{k} + 3\tilde{k}^2 + 12)/(32m) \tag{A.43}$$

Plugging this all together, we obtain

$$\int_{A_{n,j,k}} n\, t^2 g_{n,j,k}(t)\, dt = (c_{n,j,k} + \mathrm{o}(\tfrac{1}{n})) \int_{\tilde{A}_{n,j,k}} 2^{\frac{5}{2}} m^{3/2}\; t^2 f(t)\, \tilde{\varphi}(t)\, \tilde{h}(t, \tilde{\jmath}, \tilde{k}, n)\, dt$$

Substituting $t(y) = \frac{y + \tilde{x}_{n,j,k}}{\sigma_n}$, we get

$$\int_{A_{n,j,k}} n\, t^2 g_{n,j,k}(t)\, dt = \int c_{n,j,k}(1 + \tfrac{f_1}{f_0} t(y) + \tfrac{f_2}{f_0} t(y)^2 + \mathrm{o}(n^{-1}))\, \varphi(y)\, h(y, \tilde{\jmath}, \tilde{k}, n) \tfrac{(y+\tilde{x}_{n,j,k})^2}{4f_0^2}\; \mathrm{I}_{\tilde{A}_{n,j,k}}(t(y))\, dy$$

As $\tilde{x}_{n,j,k} = \mathrm{O}(n^0)$,

$$\left\{ |y + \tilde{x}_{n,j,k}| \le 2k_2 \sqrt{\log(n)}(1 + \mathrm{o}(n^0)) \right\} = \left\{ |y| \le 2k_2 \sqrt{\log(n)}(1 + \mathrm{o}(n^0)) \right\} =: A_n^0$$

For the aggregation of the factors we use MAPLE, giving

$$\int_{A_{n,j,k}} n\, t^2 g_{n,j,k}(t)\, dt = \int_{A_n^0} \left[\frac{(y + \sqrt{2}\,\tilde{\jmath})^2}{4{f_0}^2} + P_{1;n,\tilde{\jmath},\tilde{k}}(y) m^{-1/2} + P_{2;n,\tilde{\jmath},\tilde{k}}(y) m^{-1} + \mathrm{o}(n^{-1}) \right] \varphi(y)\, dy \tag{A.44}$$

for polynomials in y, $P_{1;n,\tilde{\jmath},\tilde{k}}$ and $P_{2;n,\tilde{\jmath},\tilde{k}}$ obtained by our MAPLE-procedure `getasintegrand`, where $P_{1;n,\tilde{\jmath},\tilde{k}}$ is defined as

$$\tfrac{y^2\tilde{k}(y^2-1)}{16f_0^2} + \tfrac{\sqrt{2}y^3 f_1(2-y^2)}{32f_0^4} + \left(\tfrac{\sqrt{2}y\tilde{k}(y^2+1)}{8f_0^2} + \tfrac{y^2 f_1(3-2y^2)}{8f_0^4} \right) \tilde{\jmath} + \left(\tfrac{(3+y^2)\tilde{k}}{8\,f_0^2} + \tfrac{(4-5y^2)\sqrt{2}\,f_1\,y}{16f_0^4} \right) \tilde{\jmath}^2 - \tfrac{f_1 y^2}{4f_0^4}\,\tilde{\jmath}^3 - \tfrac{f_1\sqrt{2}\,y}{4f_0^4}\,\tilde{\jmath}^4$$

and $P_{2;n,\tilde{\jmath},\tilde{k}}$ as

$$\begin{aligned}
P_{2;n,\tilde{\jmath},\tilde{k}}(y) = {} & \tfrac{\tilde{k}^2(y^6-2y^4-y^2)+y^2(7-4y^4)}{128f_0^2} + \tfrac{f_1\tilde{k}\,\sqrt{2}y^3(-2+5y^2-y^4)}{128f_0^4} + \tfrac{f_2 y^4(3-y^2)}{192f_0^5} + \tfrac{f_1^2 y^6(y^2-5)}{256f_0^6} + \\
& + \left(\tfrac{\sqrt{2}\left(\tilde{k}^2 y(9+6y^2+3y^4)+28y(3-y^4)\right)}{192\,f_0^2} + \tfrac{\tilde{k}\,f_1\,y^2\,(-2y^4+5y^2+3)}{32\,f_0^4} + \tfrac{\sqrt{2}\,y^3 f_2(12-5y^2)}{192\,f_0^5} + \tfrac{\sqrt{2}\,y^5 f_1^2(3y^2-13)}{128\,f_0^6} \right) \tilde{\jmath} + \\
& + \left(\tfrac{\tilde{k}^2(45+18y^2+3y^4)-100y^4-24y^2+84}{192f_0^2} + \tfrac{f_1\,\tilde{k}\,\sqrt{2}y(12-y^2-5y^4)}{64f_0^4} + \tfrac{y^2 f_2(9-5y^2)}{48f_0^5} + \tfrac{{f_1}^2 y^2(13y^4-41y^2-12)}{128f_0^6} \right) \tilde{\jmath}^2 + \\
& + \left(\tfrac{\sqrt{2}y(3-5y^2)}{12f_0^2} - \tfrac{\tilde{k}f_1 y^2\,(y^2+3)}{16f_0^4} + \tfrac{f_2\,\sqrt{2}y(10-9y^2)}{96f_0^5} + \tfrac{f_1^2\,\sqrt{2}y(3y^4-5y^2-4)}{32f_0^6} \right) \tilde{\jmath}^3 + \left(\tfrac{1-y^2}{4{f_0}^2} + \tfrac{f_2(1-3y^2)}{48\,{f_0}^5} + \tfrac{{f_1}^2(2y^4-1)}{32\,{f_0}^6} \right) \tilde{\jmath}^4
\end{aligned}$$

By the restriction in A_n^0, we obtain that $|y| = \mathrm{O}(\sqrt{\log(n)})$, while $\deg(P_{1;\cdot}; y) = 5$ and $\deg(P_{2;\cdot}; y) = 8$. Hence, the integrand is apparently of form $(y + \sqrt{2}\,\tilde{\jmath})^2/(4{f_0}^2) + \mathrm{O}(\sqrt{\log(n)^5/n})$, and thus, eventually in n, is maximized—up to $\mathrm{O}(\sqrt{\log(n)^5/n})$—for $|\tilde{\jmath}|$ maximal, i.e. $|\tilde{\jmath}| = \tilde{k}/2$. Even more so, if $f_1 = 0$, the $m^{-1/2}$–term, too, is maximized for $|\tilde{\jmath}| = \tilde{k}/2$. As the highest power in $P_{2;\cdot}$ occurring to y in a $\tilde{\jmath}$–term without f_1 is 4, the integrand is maximized up to $\mathrm{O}((\log(n)^4/n))$ for $|\tilde{\jmath}| = \tilde{k}/2$.
Condition $|\tilde{\jmath}| = \tilde{k}/2$ is equivalent to $j_k(t) \equiv k$ or $j_k(t) \equiv 0$. But this is the case—up to $\mathrm{o}(n^{-1})$—if condition (3.5) or (3.6) is in force, as then up to mass of order $\mathrm{o}(n^{-1})$ the contamination is either concentrated left

of $-\frac{k_2}{f_0}\sqrt{\log(n)/n}$ or right of $\frac{k_2}{f_0}\sqrt{\log(n)/n}$ for any sample with no more than $k_1 r\sqrt{n}$ contaminations. With respect to (A.31), this suffices to obtain that (II) is $o(n^{-1})$.
Later, after having integrated out y, we will see that if $f_1 = 0$, the approximation up to order n^{-1} is identical for $j_k(t) \equiv k$ and $j_k(t) \equiv 0$, whereas if $f_1 > 0$ it pays off for nature to contaminate by positive values and, correspondingly, by negative values if $f_1 < 0$. We consider $j_k(t) \equiv k$ here.
Up to $o(n^{-1})$, $\int_{A_{n,k,k}} n\, t^2 g_{n,k,k}(t)\, dt$ is

$$\int_{A_n^0} \Big[\tfrac{1}{4f_0^2}(y^2 + \tilde{k}^2/2) + Q_{1;n,\tilde{k}}(y) m^{-1/2} + Q_{2;n,\tilde{k}}(y) m^{-1} + \tilde{g}(n,k,y) \Big] \varphi(y)\, dy \tag{A.45}$$

with some skew-symmetric polynomial $\tilde{g}$ in y of degree 5 that is uniformly bounded in n on A_n^0, and for some even-symmetric polynomials $Q_{1;n,k}(y)$ and $Q_{2;n,k}(y)$; we only present the definition of $Q_{1;n,k}(y)$ below; for $Q_{2;n,k}(y)$, we refer the reader to the corresponding `MAPLE`-procedure `getasrisk`.

$$Q_{1;n,\tilde{k}}(y) = \tfrac{3\tilde{k}^3}{32f_0^2} + \left(\tfrac{f_1(\tilde{k}^3 - 3\tilde{k})}{32f_0^4} + \tfrac{\tilde{k}^3 - 2\tilde{k}}{32f_0^2} \right) y^2 + \left(\tfrac{\tilde{k}}{16f_0^2} + \tfrac{f_1\tilde{k}}{8f_0^4} \right) y^4$$

Using Lemma A.6 we see that we may drop the restriction $|y| \le 2k_2\sqrt{\log(n)}$ and integrating y out, up to $o(n^{-1})$, we get that $\int_{A_{n,k,k}} n\, t^2 g_{n,k,k}(t)\, dt$ is

$$\begin{aligned} &\tfrac{1+\tilde{k}^2/2}{4f_0^2} + \Big[(\tfrac{1}{8f_0^2} - \tfrac{3f_1}{16f_0^4})\tilde{k} + (\tfrac{1}{8f_0^2} + \tfrac{f_1}{32f_0^4})\tilde{k}^3 \Big] m^{-1/2} + \Big[(\tfrac{-1}{4f_0^2} - \tfrac{f_2}{32f_0^5} + \tfrac{15f_1^2}{128f_0^6}) + \\ &+ (\tfrac{-3}{16f_0^2} + \tfrac{3f_1}{16f_0^4} - \tfrac{f_2}{32f_0^5} + \tfrac{15f_1^2}{128f_0^6})\tilde{k}^2 + (\tfrac{3}{32f_0^2} - \tfrac{f_2}{384f_0^5} + \tfrac{3f_1}{64f_0^4} - \tfrac{5f_1^2}{512f_0^6})\tilde{k}^4 \Big] m^{-1} \end{aligned}$$

Corollary A.3 gives that we may ignore the fact that k is restricted to $k \le k_1 r\sqrt{n}$ and so with Lemma A.5, we may simply integrate out k. After substituting $n = 2m + 1$ we thus indeed get

$$\begin{aligned} \sup_{G^{(n)}} n\, [\mathrm{MSE}(\mathrm{Med}_n, G^{(n)})] = \frac{1}{4f_0^2} \Bigg\{ &(1 + r^2) + \tfrac{r}{\sqrt{n}} \Big(2(1+r^2) + \tfrac{f_1(r^2+3)}{2f_0^2} \Big) + \\ &+ \tfrac{1}{n} \Big((3r^4 + 3r^2 - 2) + \tfrac{3r^2 f_1(3+r^2)}{2f_0^2} - \tfrac{f_2(r^4+6r^2+3)}{12f_0^3} + \tfrac{5f_1^2(r^4+6r^2+3)}{16f_0^4} \Big) \Bigg\} + o(n^{-1}) \end{aligned}$$

Considering both cases $j_k(t) \equiv k$ and $j_k(t) \equiv 0$ simultaneously, we get (3.2) with (3.3) and (3.4). □

A.7 Proof of Proposition 3.4—pure quantiles and randomization

The proof for the pure quantiles is just as in the odd case and thus skipped. We only draw the attention to the different behaviour of the $1/\sqrt{n}$-correction term for positive and negative contamination which explains (3.17) in this case. For the bias corrected version M_n'', with the same techniques as in the proof of Theorem 3.2, we calculate the bias of $\sqrt{n}\, X_{[(m+1):n]}$ under F. This gives $\sqrt{n}\left|\mathrm{Bias}(X_{[m+1:n]}, F^n)\right| = B_{n,1} + B_{n,2}$ for

$$B_n = B_{n,1} + B_{n,2}, \qquad B_{n,1} = \frac{1}{2f_0\sqrt{n}}, \qquad |B_{n,2}| = \frac{|f_1|}{8f_0^3\sqrt{n}} \tag{A.46}$$

The same terms but with different signs are obtained for $\sqrt{n}\left|\mathrm{Bias}(X_{[m:n]}, F^n)\right|$. We only consider $B_{n,1}$ here, which arises no matter if we have symmetry or not and gives the bias corrected version M_n'' with the $a_{i,j}$ terms as in Proposition 3.5.

Remark A.7 We note that in all variants of the sample median up to now a minor deterministic improvement is possible if $f_1 \neq 0$, when we consider the bias-corrected estimators

$$M_n^\flat := M_n - \frac{1}{\sqrt{n}} B_{n,2} = M_n + \frac{f_1}{8f_0^3 n} \tag{A.47}$$

Except for the pure quantiles for even n, this renders all variants bias–free up to $o(n^{-1})$ in the ideal model.

A.8 Proof of Proposition 3.4—the midpoint-estimator

For the midpoint–estimator $\bar{M}_n$, we need the common law of the pure quantile estimators $X_{[m:n]}$ and $X_{[(m+1):n]}$. So more generally, we start with the common law of $(Y, Z) := (X_{[\nu_1:n]}, X_{[\nu_2:n]})$ for $1 \le \nu_1 < \nu_2 \le n$, $X_i \overset{\text{i.i.d.}}{\sim} F$, $i = 1, \ldots, n$ and $F(dx) = f(x)\,dx$, see David (1970, pp. 9–10), and in our case ($n \hat{=} 2m$, $\nu_1 \hat{=} m$, $\nu_2 \hat{=} m+1$) leads us to the density of the midpoint estimator $\bar{M}_{2m} = (Y+Z)/2$

$$g_n(t) = (2m)^2 \binom{2m-1}{m} \int_t^\infty [F(2t-u)\,(1-F(u))]^{(m-1)}\, f(u)\, f(2t-u)\, du \tag{A.48}$$

This gives for $(2m)\,\mathrm{MSE}(\bar{M}_{2m}, F)$, after substituting $s = 2t-u$, and using Fubini

$$\begin{aligned}&(2m)\,\mathrm{MSE}(\bar{M}_{2m}, F) = \\ &2m^2 \binom{2m-1}{m} \int \int_{-\infty}^{u} \frac{(u+s)^2}{4} [F(s)\,(1-F(u))]^{(m-1)}\, f(u)\, f(s)\, ds\, du\end{aligned} \tag{A.49}$$

We skip the argument showing how to choose a risk maximizing contamination. In the `MAPLE` script, however, we have detailed out a corresponding argument for $j(t)$ the number of contaminated observations larger than t. Without loss of generality, we work with the case of contamination to the right. Analogue arguments as in the preceding cases show that given we have k observations contaminated to ∞, we get as expression for the (conditional) $\mathrm{MSE}_{|K=k}$:

$$\begin{aligned}(2m)\,\mathrm{MSE}_{|K=k} \;=\; &(2m)(2m-k) \binom{2m-1-k}{m-k} \int F(u)^{(m-k)} (1-F(u))^{(m-1)}\, f(u) \times \\ &\times \int_{-\infty}^{u} \tfrac{(m-k)(u+s)^2}{4F(u)} \left(1 - \tfrac{F(u)-F(s)}{F(u)}\right)^{(m-1-k)} f(s)\, ds\, du\end{aligned} \tag{A.50}$$

which we have written in a way to be able parallel the preceding subsections. Denote the value of the inner integral by $H_k(u)$ and

$$\Delta(s,u) := (F(u) - F(s))/F(u) \tag{A.51}$$

In the inner integral, $0 \le \Delta(s,u) \le 1$, and for $\Delta(s,u) > \alpha > 0$, $H_k(u)$ will decay exponentially while being dominated, so if we introduce

$$\delta(u) := \sup\{s < u \,\big|\, F(s) \le (1-\alpha)\,F(u)\} \tag{A.52}$$

in fact we may restrict the inner integral to

$$H_k(u) = \mathrm{o}(m^{-1}) + \frac{m-k}{4F(U)} \int_{\delta(u)}^{u} (u+s)^2\, (1-\Delta(s,u))^{(m-1-k)}\, f(s)\, ds \tag{A.53}$$

But then expanding $\log(1-\Delta(s,u))$, and in order to get the right order for the expansion substituting $u = \tilde{u}/\sqrt{m}$, $s = \tilde{s}/\sqrt{m}$—according to case (I), i.e.; $|u| \le \text{const}\,\sqrt{\log(m)/m}$. Thus, for polynomials $\bar{Q}_i$ in $\tilde{s}, \tilde{u}$ defined in analogy to the Q_i in the to odd-sample case and with may be looked up in the `MAPLE` script,

$$\Delta(s,u) = 2\,\tfrac{f_0(\tilde{s}-\tilde{u})}{\sqrt{m}} + \tfrac{\bar{Q}_0(\tilde{s},\tilde{u})}{m} + \tfrac{\bar{Q}_1(\tilde{s},\tilde{u})}{m^{3/2}} + \tfrac{\bar{Q}_2(\tilde{s},\tilde{u})}{m^2} + \mathrm{O}((\tfrac{\log(n)}{n})^{5/2})$$

Hence we get

$$(m-k-1)\log(1-\Delta(s,u)) - \sqrt{m}(2\,f_0(\tilde{s}-\tilde{u})) = \texttt{logH21}(s,u) + \mathrm{o}(\sqrt{\log(n)/n}) \tag{A.54}$$

for some function `logH21`, the exact expression of which may be produced in the corresponding `MAPLE` script. Thus, denoting the term $\exp(2\sqrt{m}\,f_0(\tilde{s}-\tilde{u}))$ by $H_{;1}(s,u)$, we get

$$(1-\Delta(s,u))^{(m-k-1)} = H_{;1}(s,u)\exp(\texttt{logH21}) \times (1 + \mathrm{o}(\sqrt{\log(n)/n})),$$

Now, if we write $H_{;2,2}(s,u)$ for $(s+u)^2\, f(s)$, and $H_{k;2,2}(s,u)$ for $\exp(\texttt{logH21})$, and if we introduce $H_{k;2}(s,u) := H_{;2,1}(s,u) H_{k;2,2}(s,u)$, we get

$$4F(u)\,H_k(u) = \mathrm{o}(n^{-2}) + \int_{\delta(u)}^{u} H_{;1}(s,u)\, H_{k;2}(s,u)\, ds$$

The next step is to integrate out s where we may drop the lower restriction again due to the exponential decay far out for large values of s. After three times of integration by parts we come up with

$$4F(u)\,H_k(u) = \mathrm{o}(n^{-2}) + \sum_{i=0}^{2} \frac{(-1)^i}{(2\sqrt{m}\,f_0)^{(i+1)}} H_{;1}(s,u) \frac{\partial^i}{\partial s^i} H_{k;2}(s,u)\Big|_{-\infty}^{u} \tag{A.55}$$

that is we may restrict ourselves to these terms for our purposes. These differentiations can be done by the `MAPLE` command `diff`. Noting that essentially $t = \mathrm{O}(\sqrt{\log(n)/n})$, we hence get for the inner integral H

$$H_k(t) = t^2 + \tfrac{1}{n}[(\tfrac{f_1}{2f_0} - 1)t^2 - \tfrac{1}{2f_0}t] - \tfrac{1}{\sqrt{n^3}}\tfrac{k}{2f_0}t + \tfrac{1}{n^2}\tfrac{1}{8f_0^2} + \mathrm{o}(n^{-2})$$

So in formula (A.10) (with $j \equiv k$) we replace t^2 by $H_k(t)$ and arrive at

$$\sup_{G^{(n)}} n\,\mathrm{MSE}(\bar{M}_n, G^{(n)}) = n \sum_{k=0}^{m} \int H_k(t)\, g_{n,k,k}(t)\, dt \;\; P(K=k) + \mathrm{o}(n^{-1}) \tag{A.56}$$

Proceeding now just as in the preceding subsections, we obtain the assertion.

A.9 Proof of Proposition 3.9

For $t > \sqrt{\log(n)/n}/(2f_0)$, let

$$A_{k,t} := \Big\{ \sum_i U_i(2\,\mathrm{I}(X_i \le t) - 1) \le k-1 \Big\} \tag{A.57}$$

Hence if $t > \sqrt{\log(n)/n}/(2f_0)$, by (3.20), for all $k > (1-\delta)r\sqrt{n}$,

$$\Pr(A_{k,t} \,\Big|\, K=k) \ge p_0 \tag{A.58}$$

Now we proceed as in the proof to Theorem 3.2. But $t > \sqrt{\log(n)/n}/(2f_0) \iff y > \sqrt{\log n}$ in (A.45). Hence on the event $A_{k,t}$ for $y \in [\sqrt{\log n}; k_2\sqrt{\log n}\,)$, we get the bound $\bar{j}(t) \le (k-1)/\sqrt{n}$, while for $y \in (-k_2\sqrt{\log(n)};\, \sqrt{\log n}\,)$ respectively on ${}^cA_{k,t}$, we bound $\bar{j}(t)$ by $k/\sqrt{n}$. Integrating out these two y-domains separately, we obtain

$$\begin{aligned} & n\left(\mathrm{MSE}[\mathrm{Med}_n, G_0^{(n)} \,|K=k\,] - \mathrm{MSE}[\mathrm{Med}_n, G_\flat^{(n)} \,|K=k\,]\right) \ge \\ \ge\; & \frac{p_0}{2f_0} \int_{\sqrt{\log n}}^{k_2\sqrt{\log n}} \left(s/\sqrt{n} + \tilde{k}/\sqrt{2n} - 1/(2f_0 n)\right)\varphi(s)\,ds + \mathrm{o}(n^{-1}) \end{aligned}$$

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 $a_1 = 2\sqrt{\log n}$, $a_2 = k_2\sqrt{\log n}$, and as $\varphi(a_2) = \mathrm{o}(n^{-1})$,

$$n\left(\mathrm{MSE}[\mathrm{Med}_n, G_0^{(n)} \,|K=k\,] - \mathrm{MSE}[\mathrm{Med}_n, G_\flat^{(n)} \,|K=k\,]\right) \ge \frac{p_0}{2\sqrt{2\pi}\,nf_0} + \mathrm{o}(n^{-1})$$

By Lemma A.2, the restriction to $(1-\delta)r\sqrt{n} < K < k_1 r\sqrt{n}$ may be dropped, and we obtain the assertion. The case of an even sample size is proved similarly. □

Extra Material

On [site of the journal] we have additional supplementary material for this article: This comprises extended tables, details to points alluded to in remarks, but in particular a `MAPLE` script, referred to in the proofs.

Fig. 2 A horrifying example: (The first two pages of) the expression for (A.44) got from `MAPLE`; of course, after integration terms get much more treatable, as visible in Theorem 3.2.

Acknowledgement

Thanks are due to J. Picek for drawing the author's attention to Jurečková and Sen (1982).

References

Abramowitz M. and Stegun I.A. (Eds.) (1984): *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Selected Government Publications. National Bureau of Standards, Washington, D.C. Reprint of the 1972 edition..

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

Cabrera J., Maguluri G. and Singh K. (1994): An odd property of the sample median. *Stat. Probab. Lett.*, **19**(4): 349–354.

David H. (1970): *Order statistics.*. John Wiley & Sons, Inc.

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.

Duttweiler D. (1973): The mean-square error of Bahadur's order-statistic approximation. *Ann. Stat.*, **1**: 446–453.

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

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

Jurečková J. and Sen P. (1982): M-estimators and L-estimators of location: Uniform integrability and asymptotic risk-efficient sequential versions. *Commun. Stat., Sequential Anal.*, **1**: 27–56.

—— (1996): *Robust statistical procedures: asymptotics and interrelations.*. John Wiley & Sons Ltd.

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

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. (1994): *Robust asymptotic statistics*. Springer.

van der Vaart A. (1998): *Asymptotic statistics*. Cambridge Univ. Press.

Witting H. (1985): *Mathematische Statistik I: Parametrische Verfahren bei festem Stichprobenumfang*. B. G. Teubner, Stuttgart.

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