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)
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. 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 $\tilde{U}_n(r)$ 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

$$\sup_{G^{(n)} \in \tilde{U}_n(r)} n \{ \text{MSE}(M_n, G^{(n)}) \} = \frac{1}{4f_0^2} (1 + r^2) + \frac{r}{\sqrt{n}} a_1 + \frac{1}{2} a_2 + o(\frac{1}{n}) \quad (1.1)$$

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 $O_p(n^{-1/4})$—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.

In proving the main theorem, we use MAPLE at large to calculate tedious, lengthy asymptotic expansions which are hardly presentable in the framework of an article.

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 $\theta$, observations $X_i \sim P_\theta$ and errors $u_i$ given by

$$X_i = u_i + \theta, \quad u_i \overset{\text{iid}}{\sim} F$$

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

$$F(0) = \frac{1}{2}$$

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

$$f(x) = f_0 + f_1 x + \frac{1}{2} f_2 x^2 + O(x^{2+\delta_0}), \quad f_0 > 0$$

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

$$\int |x|^\delta f(x) \, dx < \infty$$

Remark 1.1 (a) Condition (1.3) is taken from Jurečková and Sen (1982) and is both necessary and sufficient for finiteness of $E_F[M_n]^y$ for any $y > 0$, where $M_n$ is the sample median $\text{Med}_n$ 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 $0 < \eta \leq \delta$, 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 $Q_n = Q_n(r)$ of distributions

$$Q^{(n)} := \bigotimes_{i=1}^n G_{n,i}, \quad G_{n,i} = (1 - r/\sqrt{n}) F + r/\sqrt{n} H_{n,i}$$
for arbitrary, uncontrollable contaminating distributions $H_{ij}$. As usual, we interpret $G^{(n)}$ as the distribution of the vector $(X_i)_{i \leq n}$ with components

$$X_i := (1 - U_i)X_i^a + U_i X_i^b$$

for $X_i^a, U_i, X_i^b$ stochastically independent, $X_i^a \sim F$, $U_i \sim \text{Bin}(1, r/\sqrt{n})$, and $(X_i^b) \sim H$ for some arbitrary $H \in \mathcal{M}_1(\mathbb{R}^d)$. In this setup the median can be understood as an asymptotically linear estimator with influence curve $\psi_{\text{Med}}$, allowing the expansion

$$\text{Med}_n = \frac{1}{n} \sum_{i=1}^{n} \psi_{\text{Med}}(X_i) + o_P(n^{-1/2}), \quad \psi_{\text{Med}}(x) = \frac{\text{sign}(x)}{2f_0}$$

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

$$\text{MSE}_M(S_n, G) := E_G(\min(n S_n^2, M)),$$

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

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

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 3.3 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.

Remark 1.2 It took me some time to write things up in a readable fashion: In order not to slay down the reader with vast number of terms, in the proof section, we instead describe verbally how we got them referring to a corresponding Maple script available in the internet. To give you an idea of how tedious terms become, we have included a page of Maple output on page [21] as a horrifying example.
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 \( Q_n \) 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 \( P \) be the location model from (1.2) with \( f(0) > 0 \) and let \( \text{Med}_n \) 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 \( G_n(0) = \{(1 - \frac{r}{\sqrt{n}})F + \frac{r}{\sqrt{n}} I_{\{x_0\}}\} \)

\[
\text{MSE}(\text{Med}_n, G_n(0)) > C
\]  

(2.1)

although, uniformly in \( Q_n \),

\[
\sqrt{n} \left( \text{Med}_n - \frac{1}{n} \sum_{i=1}^{n} \int \psi_{\text{med}} dG_n \right) \circ G_n \overset{w}{\rightarrow} N(0, (2f(0))^{-2})
\]  

(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 \( \tilde{Q}_n(r, \varepsilon) \) of conditional distributions

\[
G_n = \mathcal{L}[\{(1 - U_i)X_i^a + U_iX_i^b \} | \limsup_n \frac{1}{n} \sum U_i \leq \varepsilon ]
\]  

(2.3)

If we apply the Hoeffding inequality (Hoeffding (1963, Thm. 2)) to \( K = \sum U_i \) for the switching variables \( U_i \) from (1.7), we obtain

\[
P(K > m) \leq \exp \left(-2n(\varepsilon - \frac{m}{2n})^2 \right)
\]  

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

\[
P(A_n | B_n) = P(A_n \cap B_n) / P(B_n) = P(A_n)(1 + O(\varepsilon^{-2n(1+\delta)})
\]
2.3 Connection to the breakdown point

Our definition of the neighborhood $\tilde{Q}_n(r)$ 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 $\varepsilon(S_n)$ observations are contaminated. This number $\varepsilon(S_n)$ is actually just the finite sample ($\varepsilon$-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 $P = \{P_\theta, \theta \in \Theta\}$ be a parametric model and $X^n$ be $\mathbb{R}^k$-valued observations distributed according to the ideal situation $P_\theta$. 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 $\ell \geq 0$ on some (thinned out) neighborhood or not. To this end we define $\tilde{Q}_n(r, \varepsilon)$ analogously to (2.3). Assume that there is some $\bar{\varepsilon} > 0$ such that for each $n \in \mathbb{N}$ and $k \leq \bar{k} := \lfloor \frac{n}{\bar{\varepsilon}} \rfloor - 1$

$$\varepsilon(S_n) := \inf \{\varepsilon^*(X_{n-k}, S_n) \mid X_{n-k} = (x_1, \ldots, x_{n-k}) \text{ a possible sample configuration, } k \leq \bar{k} \geq \bar{\varepsilon} > 0 \} \quad (2.5)$$

where $\varepsilon^*(X, S)$ is the finite sample ($\varepsilon$-contamination) breakdown point of $S$ at sample $X$. Then, by an analogue argument to that of Proposition 2.1, we have uniform convergence of the risk holds on $\tilde{Q}_n(r, \varepsilon)$ for any $\varepsilon < \bar{\varepsilon}$. Note that this thinning out for continuous loss functions $\ell$ is not needed if $\ell$ is bounded.

3 Higher order asymptotics for the MSE of the sample median

For $H \in M_1(\mathbb{E}^k)$ 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 $\kappa_n$ in $\mathbb{R}$ in $(0, \infty)$, and in $[1, \ldots, n]$, respectively. We say that the sequence $(H^{(i)}) \subset M_1(\mathbb{E}^k)$ is $\kappa_n$-concentrated left [right] of $c_n$ up to $o(d_n)$, if for each sequence of ordered sets $I_n$ of cardinality $i_n \leq \kappa_n$

$$1 - H^{(i)}((-\infty; c_n]^k) = o(d_n) \quad \left[ 1 - H^{(i)}((c_n, \infty)^k) = o(d_n) \right] \quad (3.1)$$
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 $\widetilde{Q}(r, \varepsilon)$ of (2.3) it holds

$$\sup_{G^{(n)}} \text{MSE}(\text{Med}_n, G^{(n)}) \leq \frac{1}{4f_0^2} a_1 + \frac{r}{\sqrt{n}} a_2 + o(1/n)$$

(3.2)

for

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

(3.3)

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

(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 $\kappa_n = r k_1 r \sqrt{n}^{-1}$, eventually in $n$, either

$$(H_n) \text{ is } \kappa_n-\text{concentrated left of } -\frac{k_1}{\kappa_n} \sqrt{\log(n)/n} \text{ up to } o(n^{-1})$$

(3.5)

or

$$(H_n) \text{ is } \kappa_n-\text{concentrated right of } \frac{k_1}{\kappa_n} \sqrt{\log(n)/n} \text{ up to } o(n^{-1})$$

(3.6)

More precisely, if $f_1 < 0$ and $f_0 > 0$, the maximal MSE is achieved up to $O(n^{-2})$ by contaminations according to (3.5) or (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 $\text{MSE}$ up to order $o(n^{-1/2})$, one may weaken assumption (1.4) to: For some $\delta > 0$,

$$f(s) = f_0 + f_1 s + O(s^{1+\delta}), \quad f_0 > 0$$

(3.7)

(c) Conditions (3.5) and (3.7) 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 $\text{MSE}$ — an astonishingly modest contamination! With respect to (1.3), this is plausible however, as $|\text{Med}_n|$ attains its maximal value for any $s \neq 0$.

The thinning out of the marginals by means of Definition 3.1 even tells us that of the $n$ potentially contaminating $X_i^n$ only all subsets of cardinality roughly $\sqrt{n}$ need to be “large” at all, the remaining coset of cardinality order $n(1 + o(1))$ of contaminations might even stem from the ideal situation!

As shown in Propositions 3.4 and 3.6 resp. (3.7), (3.8) 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 $n \text{MSE}$— 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 $\text{MSE}$ 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 $U \sim \text{Bin}(1, 1/2)$
- the midpoint-estimator $\bar{M}_n := (X_{[m:n]} + X_{[(m+1):n]})/2$
- the bias corrected estimator $M''_n := (X_{[m:n]} + \frac{1}{\sqrt{n}})$

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 $Q_{n}(r, \varepsilon)$ of (3.3) it holds

$$\sup_{G^{(n)}} n \text{MSE}(M_n, G^{(n)}) = \frac{1}{4n^2} \left(1 + 3\rho^2\right) + \frac{\varepsilon}{4n} \left(a_{1,0} + a_{1,1} c_{1,0}\right) + \frac{\varepsilon^2}{2n} \left(a_{2,0} + a_{2,1} c_{1,0} + a_{2,2} r + a_{2,2} c_{1,0}\right) + \alpha(\frac{1}{n})$$

(3.8)

for some real numbers $a_{ij} = a_{ij}(M_n)$ 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 $a_{ij}$.] Splitting up $a_{2,0}$, $a_{2,1}$, $a_{2,2}$ according to

$$a_{2,0} = a_{2,0,c} + a_{2,0,r}, \quad a_{2,1} = a_{2,1,c} + a_{2,1,r}, \quad a_{2,2} = a_{2,2,c} + a_{2,2,r}$$

(3.9)

we get

(a) Identical terms for all variants:

$$a_{2,0,c} = \frac{5(r^4 + 6r^2 + 3)}{16}, \quad a_{2,1,c} = -1/4, \quad a_{2,2,c} = -\frac{(r^4 + 6r^2)}{12}$$

(b) Varying terms in the ideal model:

$$a_{2,0,r}(M'_n) = -2, \quad a_{2,0,c}(\bar{M}_n) = -3, \quad a_{2,0,c}(X_{[m:n]}) = a_{2,0,c}(X_{[(m+1):n]}) = a_{2,0,c}(M''_n) = -1$$

(c) Remaining $a_{ij}$ for $\bar{M}_n$, $M'_n$, and $M''_n$:

$$a_{1,0}(M'_n) = a_{1,0}(M''_n) = a_{1,0}(\bar{M}_n) = 2(1 + r^2), \quad a_{1,1}(M'_n) = a_{1,1}(M''_n) = a_{1,1}(\bar{M}_n) = (r^2 + 3) \text{sign}(f_1)/2, \quad a_{2,0,r}(M'_n) = a_{2,0,r}(M''_n) = 3r^4 + 3r^2 = a_{2,0,r}(M'_n) - 2r^2 \text{sign}(f_1)$$

$$a_{2,1,c}(M'_n) = a_{2,1,c}(M''_n) = 0, \quad a_{2,1,c}(\bar{M}_n) = 1, \quad a_{2,1,r}(M'_n) = a_{2,1,r}(M''_n) = \frac{3r^3(3r^2 + 6r + 4)}{2} \text{sign}(f_1) = a_{2,1,r}(M''_n) - r^2$$

(3.12)

(3.13)

(3.14)
Correspondingly, let
\[ a_{2,1,\varepsilon}(X_{[m:n]}) = 3/2 = -a_{2,1,\varepsilon}(X_{[m:n]}) \]  
(3.15)

For \( X_{[m:n]} \) and \( X_{([m+1]:n]} \), condition \((3.5)\) / \((3.6)\) applies if \( 4f_0^2 > \lceil \cdot \rceil - (3 + r^2)f_1 \).

Correspondingly, let
\[ s' = \begin{cases} 
1 & \text{for } X_{[m:n]} \\
-1 & \text{for } X_{([m+1]:n]} 
\end{cases} \]
(3.16)

and
\[ s = \text{sign}(3 + r^2)f_1 + s'4f_0^2 \]  
(3.17)

Then the remaining \( a_{i,j} \) for \( X_{[m:n]} \) and \( X_{([m+1]:n]} \) are given by
\[ a_{1,0} = 2 + 2s's + 2r^2, \quad a_{2,1,\varepsilon} = 3s's(3 + s)r^2 + r^4)/2 \]
\[ a_{2,0,\varepsilon} = 3r^4 + (3 + 4s)r^2 \]  
(3.18)

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

Remark 3.6 In case of the sample median for odd sample size,
\[ a_{1,0} = 2(1 + r^3), \quad a_{1,1} = \frac{2(1 + r^3)\text{sign}(f_1)}{2}, \quad a_{2,0,\varepsilon} = -2, \quad a_{2,1,\varepsilon} = 3s's(3 + s)r^2 + r^4)/2 \]
\[ a_{2,2,\varepsilon} = -\frac{3}{4}, \quad a_{2,3,\varepsilon} = -\frac{3}{16}, \quad a_{2,3} = \frac{3}{16} \]  
(3.19)

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, \( M_n \) up to \( o(1/n) \) has a smaller \( \text{MSE} \) than \( M_c \) — the only difference arising in term \( a_{2,0,\varepsilon} \).

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
\[ \text{MSE}(M_2, F) - \text{MSE}(M_{2n}, F) = -\frac{1}{16n^3} + o(n^{-3}) \]  
(3.20)

that is, evaluating the sample median at one more observation (from 2m to 2m + 1) deteriorates \( \text{MSE} \)! As our expansion already stops at \( o(n^{-3}) \), we cannot reproduce (3.19) to the given exactitude by means of our representations (3.2) and (3.3).

After correcting (minor) typing errors in formulae (2.2), (2.5), and (2.6) in the cited reference, we obtain (3.2) and (3.3) 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 \( \delta_0 \). Assume that, for \( K = \sum_{i=1}^{n} U_i \) and \( k > (1-\delta)r \sqrt{n} \),
\[ \Pr\left( \sum_{i=1}^{n} U_i I(X_i^0 \leq \sqrt{\log(n)}/n/(2f_0)) \geq 1 \mid K = k \right) \geq p_0 > 0 \]  
(3.20)

Then, eventually in \( n \), no such sequence of contaminations \( G^{(n)}_y \in \tilde{Q}(r) \), can attain the maximal \( \text{MSE} \) in (3.2) as in condition (3.6) (i.e. with positive bias). More precisely,
\[ \sup_{G^{(n)}_y} n \left[ \text{MSE}(M_n, G^{(n)}_y) - n \left[ \text{MSE}(M_n, G^{(n)}_y) \right] \right] \geq \frac{p_0}{2n^2f_0} + o(1/n) \]  
(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 spelled out in (3.5), (3.6), leading to \(G^{(n)}_0\) in (2.3), it holds

\[
n \cdot \text{Var}(M_n, G^{(n)}_0) = \frac{1}{4f_0} \left\{ 1 + \frac{r}{\sqrt{n}} \left( 2 + |a_0| \right) + \right. \\
+ \frac{1}{\pi} \left[ 3(f_0^2 - 3(\pi - 1)r)/2 + \frac{3(f_0^2 + 3)}{4f_0^2} + \frac{f_0^2(8r^2 + 7)}{8f_0^2} \right] + o\left( \frac{1}{n} \right) \tag{3.22} \]

\[
\sqrt{n} \cdot \text{Bias}(M_n, G^{(n)}_0) = \frac{1}{2f_0} \left\{ r^2 + \frac{1}{\sqrt{n}} \left( f_0^2 - \frac{1}{2f_0} \right) \right. \\
+ \frac{f_0^2}{2r} \left( r^2 + \frac{1}{f_0^2} \right) + \left. \frac{f_0^2(8r^2 + 7)}{8f_0^2} \right] + o\left( \frac{1}{n} \right) \tag{3.23} \]

\[
n \cdot \text{Bias}^2(M_n, G^{(n)}_0) = \frac{1}{4f_0} \left\{ r^2 + \frac{1}{\sqrt{n}} \left( 2r^2 + \frac{1}{2f_0} \right) + \right. \\
+ \frac{1}{\pi} \left[ 4r^2 + \frac{3(f_0^2 + 3)}{2f_0^2} - \frac{f_0^2(8r^2 + 7)}{16f_0^2} \right] + o\left( \frac{1}{n} \right) \tag{3.24} \]

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

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

\[
\sup_{G^{(n)}} n \cdot \text{MSE}(M_n, G^{(n)}) = \frac{\pi}{2} \left( 1 + r^2 + \frac{1}{\sqrt{n}} \left( a_0 + \frac{1}{\pi} a_0 + \frac{1}{\pi} (a_2 + 2\pi a_2) \right) \right) + o\left( \frac{1}{n} \right) \tag{3.25} \]

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

\[
n \cdot \text{MSE}(\text{Med}_n, F) = \frac{\pi}{2} \left[ 1 + \left( \frac{\pi}{2} + a_2 + \frac{1}{\pi} a_2 \right) / n \right] + o\left( \frac{1}{n} \right). \tag{3.26} \]

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

\[
n \cdot \text{MSE}(\text{Med}_n, F) \simeq o\left( \frac{1}{n} \right) + \begin{cases} 
1.5708(1 - 0.4292/n) & \text{for Med}_n, \ M_n' \\
1.5708(1 + 0.5708/n) & \text{for } X_{[\text{med}]}, X_{[\text{med} + 1/n]} \\
1.5708(1 - 1.4292/n) & \text{for } \bar{M}_n.
\end{cases} \tag{3.27} \]

This means: We overestimate \(\text{MSE}(\text{Med}_n, F)\) by the first order asympotics 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 \(X_{[\text{med}]}\), \(X_{[\text{med} + 1/n]}\) - \(M_n''\) however is underestimated.

### 4 Illustration of the results

To illustrate the approximation, we consider the case of \(F = N(0, 1)\) with a number of numerical evaluations and a small simulation study.
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 gₙ for the sample median for odd sample size from (A.5) and gₙ for the midpoint estimator for even sample size from (A.48). For the numerical calculations, we have used \textit{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 \textit{R} 2.11.0, compare \textit{R Development Core Team} (2010), we simulated \( M = 10000 \) runs of sample size \( n = 5, 10, 30, 100 \) in the ideal location model \( \vartheta = \mathcal{N}(\theta, 1) \) at \( \theta = 0 \). In a contaminated situation, we used observations stemming from

\[
G_s^{(n)} = \mathcal{L}\{[(1 - U_i)X_i + U_iX_i^{\text{cont}}] : \sum U_i \leq r/n/2^n - 1 \}
\]

for \( U_i \sim \text{Bin}(1, r/\sqrt{n}), X_i^{\text{id}} \sim \mathcal{N}(0, 1), X_i^{\text{cont}} \sim I_{100} \) 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 \( G_{th} \).

| \( n \) | exact Var \( \hat{\theta} \) | error of asymptotics |
|-----|----------------|-------------------|
|     | 1st/2nd order | 3rd order         |
|     | absolute      | relative          |
| 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 % |
| 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 % |
| 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 % |
| 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
The simulated results for $n \text{MSE}(\text{Med}_n, G_{n}^{(r)})$ come with an asymptotic 95%--confidence interval, which is based on the CLT for the variable

$$\text{empMSE}_n = \frac{n}{\text{Max MSE}} \sum_j (\text{Med}_n(\text{sample}_j))^2$$  \hspace{1cm} (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, $g_{n,k}$ 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.

### 4.2.2 Results

The simulated results for $n \text{MSE}(\text{Med}_n, G_{n}^{(r)})$ show an excellent approximation quality of our formulas (4.2) and (5.8) with specifications (5.9) to (5.14) in the ideal model.
Fig. 1 The mapping $n \mapsto \text{rel.error}(\text{maxMSE}(\text{Med}_n))$ 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 [I] 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 \( \text{maxMSE}(\text{Med}_n) \) on \( \hat{Q}_r(r) \) 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 1.1 (a)

Let \( n = 2m + 1 \) and \( r \in (0, 1). \) Necessity: With \( F = 1 - F, \) 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 \( a = \frac{1}{m+1} \)

\[
\text{Var} |\tilde{\text{Med}}_n|^p = m^2 \int_{-T}^T |F(x)|^p F(x)^m F(dx) \geq \nonumber
\]

\[
K \max \left( \int_{-T}^{\infty} x^{3-1} F(-x)^{p+1} \, dx, \int_{-T}^{\infty} x^{3-1} F(x)^{p+1} \, dx \right) \geq K \left( \int_{-T}^{\infty} |F(x)|^p F(dx) \right)^{m+1}
\]

Sufficiency: Under condition (1.5), for \( g_d(i) = |i| \delta F(i \tilde{F}(i)) \) and \( \tilde{g}_d := \sup_{i} g_d(i) \) it holds —cf. [Jurečková and Sen (1982)](2.37)

\[
\tilde{g}_d < \infty, \lim_{|i| \to \infty} g_d(i) = 0, \quad \tilde{d}_b := \int |F(i \tilde{F}(i)|^b \, dt < \infty \quad \forall b \geq 1/\delta \quad (A.1)
\]

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

\[
\text{Var} |\tilde{\text{Med}}_n|^p = m^2 \int_{-T}^T x^{3-1} \left[ F^{m+1}(x) F^m(x) + F^m(-x) F^{m+1}(-x) \right] \, dx \leq \nonumber
\]

\[
\leq m \int_{-T}^T \tilde{g}_d(x)^{1/\delta} \int_{-\infty}^{\infty} \left[ F(x) F(x)^b \right] \, dx \leq m \int_{-T}^T \tilde{g}_d(x)^{1/\delta} \tilde{d}_b < \infty
\]

The arguments for even sample size are similar. \( \square \)

#### A.2 Proof of Proposition 2.1

The assertion for uniform normality follows along the lines of [Kochetov 1994] Theorem 6.2.8: Although the assumed uniform Lipschitz continuity of the scores \( \varphi \)(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 \( d \tilde{L}(\theta) = \tilde{l}_k \) on p. 235, which in our situation is the case anyway.

Assertion (2.1) is shown by a breakdown-point argument: We interpret \( G^{(1)} \) according to (1.7), where for this proof \( X^n \triangleq I_{\hat{U}_0} \). We observe that \( \text{Med}_n \geq x_0 \) surely under \( G^{(1)} \) as soon as \( K = \sum U_i \), the number of observations stemming from \( I_{\hat{U}_0} \), is larger than \( m \). But, \( K \) being a binomial variable, the event \( |K - m| \) carries positive probability \( p_s \). So setting \( \tilde{x}_0 := \sqrt{C/p_s}, \) we get

\[
\text{MSE}(\text{Med}_n, G^{(1)}) = E_{G^{(1)}}(\text{Med}_n) \geq E_{G^{(0)}}(\text{Med}_n I_{K=m}) \geq \tilde{x}_0^2 p_s = C \quad \square
\]
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

\[
  j = j_k(t) := \#(\tilde{x}_i : \tilde{x}_i \geq t)
\]

(A.2)

In this situation we will derive the (conditional) probability that the (unique) median \( \text{Med}_n \) 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

| \( K \leq k_1 \sqrt{n} \) | \( k_1 \sqrt{n} < K \leq \rho n \) | \( K > \rho n \) |
|---------------------|-------------------------------|-------------------|
| (I)                 | (II)                          | (III) excluded    |
| \( k_2 \sqrt{\log(n)/n} \leq \| \leq k_2 \sqrt{\log(n)/n} \) | \( \| \leq k_2 \sqrt{\log(n)/n} \) |

For cases (II) to (IV), we will show that they contribute only terms of order \( o(n^{-1}) \) to \( n \text{MSE}(\text{Med}_n) \) 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 \( -k_2 f_0 \sqrt{\log(n)/n} \) or larger than \( k_2 f_0 \sqrt{\log(n)/n} \). Integrating out first \( t \) and then \( k \) we obtain the result.

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

A.4.1 Ideal Situation

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

\[
  P(X_{[k:n]} \leq t) = \sum_{l=0}^{k-1} \binom{n}{l} P(t)^l (1 - P(t))^{n-l}
\]

(A.3)

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

\[
  g(t) = np(\binom{n-1}{k-1}) P(t)^{k-1} (1 - P(t))^{n-k}
\]

(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

\[
  g_n(t) = nf(t) \left( \frac{2m}{m} \right)^{m} F(t)^m (1 - F(t))^m
\]

(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 \( Y_i := I_{X_i \leq t} \). Then the following events are identical

\[
  \{X_{[k:n]} \leq t \} = \{\bigcup_{i=1}^{n} Y_i \geq k \} = \left\{ \sum_{i=1}^{n} Y_i \geq k \right\}
\]

(A.6)

The fact that \( Y_i \overset{i.i.d.}{\sim} \text{Bin}(1, P(t)) \) entails [A.3], [A.4] follows by simple differentiating, [A.5] by plugging in \( k = m + 1 \). \( \square \)
Thus abbreviating again $j$ from the Stirling formula to be found e.g. in Abramowitz and Stegun (1984, 6.1.37):

$$A.5.1 \text{ Stirling approximations}$$

we may suggestively write

$$\text{knowing}$$

$$S$$

may suggestively write

$$S_n = S^a_{n-k} + S^b_{n}$$

giving

$$(\text{Med}_n \leq t) = \bigcup_{n=1}^{m}(S^a_{n-k} + S^b_{n}) \cap \{K = k\}$$

Splitting up again this event by the realizations of $S^b_{n}$, we get

$$(\text{Med}_n \leq t) = \bigcup_{m=1}^{m}(\omega^m_{j}(S^a_{n-k} + S^b_{n}) \leq m) \cap \{S^b_{n} = j\} \cap \{K = k\}$$

Thus, for the moment, we may consider the situation that exactly $k$ observations, $0 \leq k \leq m$, are contaminated, and exactly $j = j(t)$ of the contaminated observations are larger than $t$ and denote that event with $D_{j,k}$. As $X_{(m-j-k+1)} \leq t$ is independent from $D_{j,k}$ with $P = 1 - P$, the conditional density of $\text{Med}_n$ knowing $D_{j,k}$ is

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

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

$$n \text{MSE}(\text{Med}_n, G^{(n)}) = n \sum_{k=1}^{m} \sum_{j=1}^{m} \int \xi \cdot g_{n,j,k}(t) \, dt \cdot P(S^b_{n} = j) \cdot P(K = k)$$

### 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):

$$\binom{2m-k}{n-j} = (\frac{2m-k}{n-k})^{n-j/2} \frac{2m-k}{n-k}^{n-j-k} \frac{1}{\sqrt{2\pi(n-k)}} (1 + \rho_{n,j,k}), \quad \text{for } \frac{1}{2} - \frac{1}{m} \leq \rho_{n,j,k} \leq \frac{1}{2n}$$

$$= (\frac{2m-k}{n-j})^{n-j/2} (\frac{2m-k}{2\pi(n-j-k)^{3/2}})(1 - \frac{1}{m} + o(\frac{1}{n})), \quad \text{for } j,k = O(\sqrt{n})$$

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

**Lemma A.2** Let

$$\kappa := k_1 \log k_1 + 1 - k_1$$

Then it holds that

$$\text{Pr}(\text{Bin}(n, r/\sqrt{n}) > kr/\sqrt{n}) \leq \exp(-kr \sqrt{n} + o(\sqrt{n}))$$

**Proof** We first note that $\kappa > 0$, as $\log(x^x) > 0$ for $x > 1$ and $\kappa = \int_{1}^{\infty} \log(x) \, dx$. By Hoeffding’s inequality (Hoeffding [1963] Thm. 1, inequality (2.1)), we have for $\xi_i, i = 1, \ldots, n$ i.i.d. real-valued random variables, $|\xi| \leq M, \mu = E(\xi)$ and $0 < \epsilon < 1 - \mu$

$$P\left(\frac{1}{n} \sum_{i=1}^{n} \xi_i - \mu \geq \epsilon\right) \leq \left(\frac{\mu + \epsilon}{\mu - \epsilon}\right)^{\frac{n}{2}}$$

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

$$\Pr(B_n > k_1r\sqrt{n}) \leq \exp\left( -k_1r\sqrt{n} \log(k_1) + (n - k_1r\sqrt{n})(\log(1 - \frac{r}{\sqrt{n}}) - \log(1 - \frac{r}{\sqrt{n}})) \right)$$

For $x \in (0, 1)$, $-\frac{r}{\sqrt{n}} \leq \log(1 - x) \leq -x$. Thus the difference of the logarithms is smaller than $(k_1r)/\sqrt{n} - k_1r/\sqrt{n}$ and

$$\Pr(B_n > k_1r\sqrt{n}) \leq \exp(-x\sqrt{n} + o(\sqrt{n}))$$

$\square$

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

$$E[X^i \mid X \geq r\sqrt{n}] = o(n^{-1})$$

(A.16)

**Proof** $E[X^i \mid X \geq r\sqrt{n}] \lesssim const n^i \exp(-x\sqrt{n})$ $\square$

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

$$\left| \frac{m-j}{n} - F(t) \right| \leq k_2 \sqrt{\frac{\log n}{n}} (1 + o(n^0)) \implies |t| \leq k_2 \sqrt{\frac{\log n}{n}} (1 + o(n^0))$$

(A.17)

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

$$\frac{m-j}{n} = 1/2 + \frac{k_2j}{n} + \frac{k_2(k-2j)}{n} + o(n^{-1})$$

By (129), (123), $F(t) = 1/2 + j_0t + o(t)$; thus $\left| \frac{m-j}{n} - F(t) \right| = o(\frac{1}{\sqrt{n}}) - j_0t$. $\square$

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

- $E[X] = rn^{1/2}$, $E[X^2] = r^2n + r n^{1/2} - r^2$,

(A.19)

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

(A.20)

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

(A.21)

**Proof** Cf. the MAPLE-procedure Binmoment on the web-page. $\square$

Finally, we note the following Lemma for $N(0, 1)$ variables

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

$$E[X^k \mid X \geq c\sqrt{\log(n)}] = o(n^{-1})$$

(A.22)

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

$$E[X^k \mid X \geq c\sqrt{\log(n)}] = \begin{cases} P_k(c)\varphi(c)\sqrt{\log(n)} & \text{for } k \text{ odd} \\ P_k(c)\varphi(c) + \left[ \prod_{i=0}^{k/2-1} (2i+1) \Phi(x) \right]_{\sqrt{\log(n)}}^\infty & \text{for } k \text{ even} \end{cases}$$

for some polynomial $P_k$ of degree $k - 1$. The assertion follows, as $\varphi(c\sqrt{\log(n)}) = \varphi(0)n^{-c^2/2} = \varphi(0)n^{-(c^2/2)}$ for some $\delta > 0$, and because for the $\Phi(x)$-term, $\overline{\Phi}(x) \leq \varphi(x)/x$ for $x > 0$. $\square$
A.6 Proof for odd sample size

We recall the density \( g_{n,jk} \) from (A.9)

\[
g_{n,jk}(t) := (n-k) \left( \frac{2m-k}{m-j} \right)^{m-j} F(t)^{m-j} F(t)^{m+j-k} f(t)
\]

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

\[
\left( \frac{2m-k}{m-j} \right)^{m-j} \left( \frac{2m-k}{m+k-j} \right)^{m+k-j} \gamma_{n,jk}
\]

where

\[
\gamma_{n,jk} := \sqrt{\frac{2m-k}{m+k-j}} \frac{1}{(1 + \rho_{n,jk})}
\]

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

\[
F(t)^{m-j} F(t)^{m+j-k} \leq \left( \frac{m-j}{2m-k} \right)^{m-j} \left( \frac{m+j-k}{2m-k} \right)^{m+j-k}
\]

with equality if \( t = x_{n,jk} \) for

\[
x_{n,jk} := F^{-1}(\frac{m-j}{2m-k})
\]

Introducing

\[
\Delta F_{n,jk} := F(t) - \frac{m-j}{2m-k} \Rightarrow F(t) - F(x_{n,jk}),
\]

we see that

\[
g_{n,jk}(t) = (n-k)\gamma_{n,jk} f(t) [1 + \frac{2m-k}{m-j} \Delta F_{n,jk}]^{m-j} [1 + \frac{2m-k}{m+j-k} \Delta F_{n,jk}]^{m+j-k}
\]

**Case (III):** For \( k \in [k_1 r \sqrt{n}, \infty] \), \( 0 \leq j \leq k \), we partition the terms according to (A.29) and see that on \( |t| \leq n^2 \), the integral \( n^2 g_{n,jk}(t) \) multiplied by \( n^{-3} \) is \( o(n^3) \) 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 \( \Pr(K \geq k_1 r \sqrt{n}) \) 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 \( k \leq k_1 r \sqrt{n} \) and \( |t| > \frac{1}{n} \sqrt{\log(n)/n} \), or equivalently by Lemma [A.4]

\[
|\Delta F_{n,jk}| > k_2 \sqrt{\log(n)/n}
\]

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

\[
g_{n,jk}(t) / f(t) \leq (n-k)\gamma_{n,jk} \exp \left( \frac{2m-k}{m-j} \Delta F_{n,jk}^{m-j} \right) \leq (n-k) \sqrt{\frac{1}{m+k-j} (1 + \frac{1}{m})} \exp \left( -k_2^2 \log(m) \right)
\]

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

\[
g_{n,jk}(t) / f(t) \leq \text{const} m^{1/4} \left( 1 + o(n^{-3}) \right) = o(n^{-2})
\]

where the last equality is a consequence of \( k_2 > \sqrt{3/2} \). 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 (A.18), for \( n \) large enough, we obtain bound \( g_{n,jk} \leq n^2 f(t)(F(t)\log t)^{-1/2} \). Let \( \eta = 1/2 - c \), \( b = 2/\delta \) and \( \delta \in (0,1) \). By choosing \( n \) large enough, we may achieve that \( F(n^2) = A_n < 2^{-1/\eta} \) and \( F(n^2)^{2b} > 1 - \delta' \). So by (A.17), we get eventually in \( n \) and for some constant \( c \) and any \( \eta' > 0 \), and \( g_{n,jk} \) from the proof of Remark [1.1.4(a)]

\[
\int_{n^2}^{\infty} n^2 g_{n,jk}(t) dt \leq n^2 \frac{\log n^2}{1-\delta^2} \int_{n^2}^{\infty} \left[ (F(t)^{2b})^2 (F(t)/\log t)^{1/2} \right] dt \left( \log n \right)^{1/2-2b} \leq \frac{\delta^2 n^2}{(1-\delta')^{1/2+2b}} \left( 2L_0 \right)^{\eta'} c \exp(\eta^2 n \log n - \log \left( \frac{1}{\log A_n} - \frac{1+\delta}{2\delta} \right)) \leq \exp(-\eta' n)
\]
Case (I): Here we restrict ourselves to the case that
\[ k \leq k_1 \sqrt{n}, \quad \frac{m-j}{\sqrt{m}} - F(t) \leq k_2 \frac{\log(n)}{n} \]  
(A.32)

Doing so, we set \( u := t - x_{n,jk} \). As on (I), \( k = O(\sqrt{n}) \) as well as \( j \), we make this magnitude explicit to \textsc{maple} in the function \( \text{transf} \) by introducing the bounded variables
\[ \tilde{k} := k / \sqrt{m} \quad \text{and} \quad \tilde{j} := (k/2 - j)/\sqrt{m} \]  
(A.33)

This gives the expansion in powers of \( m^{-1/2} \)
\[ (m-j)/(2m-k) = 1/2 + \tilde{j}/(4\sqrt{m}) + \tilde{k}\tilde{j}(8m) + o(n^{-1}) \]  
(A.34)

Thus, to get an approximation to \( x_{n,jk} = F^{-1}\left( \frac{m-j}{\sqrt{m}} \right) \), we expand this in a Taylor series in powers of \( m^{-1/2} \) (compare our \textsc{maple}-procedure \texttt{asquantile}) which gives
\[ x_{n,jk} = \frac{f_0}{2} \sum_{\ell=0}^{\infty} \frac{j^{2\ell+1}}{(2\ell+1)!} \frac{(2\ell+1)!}{(2\ell)!} \left( \frac{1}{2m} \right)^{2\ell} \left( \frac{1}{2m} \right)^{2\ell} + o(n^{-1/2}) \]  
(A.35)

Furthermore,
\[ f(x_{n,jk}) = f_0 + f_1(u^2/2 + u x_{n,jk}) + f_2(u^3/6 + (u^2 x_{n,jk} + u x_{n,jk}^2)/2) + o(n^{-3/2}) \]

which implies that in (I), by (A.32), \( u \) lies in a shrinking compact, as
\[ u = F^{-1}(m-j) - F^{-1}(F(x_{n,jk}) - F(0)) = \frac{m-j}{\sqrt{m}} + o(\sqrt{m}) = O(\sqrt{m}). \]

Setting \( \Delta f_{n,jk} := F(t) - F(x_{n,jk}) \), and expanding this in a Taylor series around 0, we get
\[ \Delta f_{n,jk} = f_0 u + f_1(u^2/2 + u x_{n,jk}) + f_2(u^3/6 + (u^2 x_{n,jk} + u x_{n,jk}^2)/2) + o(n^{-3/2}) \]

and
\[ f(t) = f_0 + f_1(u + x_{n,jk}) + f_2((u + x_{n,jk})^2/2 + o(n^{-1}) \]

We turn to the constant factors now; up to now, the terms arising by applications of the Stirling formulas of subsection \texttt{A.5.1} come with \( k \)-terms in the denominators. 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 
\[ \text{asympt} \] and \texttt{collect} in \textsc{maple} (compare our function \texttt{asbinom}):
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 $\bar{\gamma}_{n, jk} := x_{n, jk}\sigma_n$, and, with $\varphi$ the density of $N(0, 1)$, use the abbreviations

$$\bar{\varphi}(t) = \varphi \circ y \circ u(t), \quad \bar{h}(t, j, k, n) = h(y \circ u(t), j, k, n) \quad (A.41)$$

We also introduce the integration domains

$$A_{n, jk} = \{ t \in \mathbb{R} \mid \frac{|\varphi(t)|}{\varphi(t)} - F(t) \leq k_2 \frac{\log(n)}{\sqrt{n}} \}, \quad \bar{A}_{n, jk} = \{ t \leq k_2 \frac{\log(n)}{\sqrt{n}} (1 + o(n^2)/f_0) \} \quad (A.42)$$

Finally, applying (A.12) and (A.30), we derive an integration constant $c_{n, jk}$ from $\gamma_{n, jk}$ from (A.24):

$$c_{n, jk} := 2 \cdot 2^{m-2} m_{\gamma} \gamma_{n, jk} = 1 - \frac{1}{2} (4m^{1/2}) + (16 \frac{32}{2} + 32 + (16 \frac{32}{2} + 32)) / (32m) \quad (A.43)$$

Plugging this all together, we obtain

$$\int_{A_{n, jk}} n^2 \bar{r}_{n, jk}(t) dt = (c_{n, jk} + o(\frac{1}{n^{3/2}})) \int_{A_{n, jk}} 2 \frac{r^2}{2} \bar{r} f(t) \bar{\varphi}(t) \bar{h}(t, j, k, n) dt$$

Substituting $t(y) = \frac{\sqrt{5}}{\sqrt{2}} \frac{\sigma_n}{\sqrt{f_0^2}}$, we get

$$\int_{A_{n, jk}} n^2 \bar{r}_{n, jk}(t) dt = \int c_{n, jk}(1 + \frac{\sqrt{5}}{\sqrt{2}} t + \frac{\sqrt{5}}{\sqrt{2}} t^2 + o(n^{-1})) \varphi(y) h(y, j, k, n) \frac{(2 \frac{\sqrt{5}}{\sqrt{2}} \frac{\sigma_n}{\sqrt{f_0^2}})}{n^{3/2}} 1_{A_{n, jk}} (t(y)) dy$$

As $\bar{\gamma}_{n, jk} = O(n^2)$, we get

$$|y + \bar{\gamma}_{n, jk}| \leq 2k_2 \frac{\log(n)}{\sqrt{n}} (1 + o(n^2)) = \{ |y| \leq 2k_2 \frac{\log(n)}{\sqrt{n}} (1 + o(n^2)) \} = A_{0, n}^\prime$$

For the aggregation of the factors we use MAPLE, giving

$$\int_{A_{n, jk}} n^2 \bar{r}_{n, jk}(t) dt = \int_0^1 \left( \left( y + \frac{\sqrt{5}}{\sqrt{2}} f \right)^2 \right) + P^1_{1, n, jk}(y)m^{-1/2} + P^2_{2, n, jk}(y)m^{-1} + o(n^{-1}) \right) \varphi(y) dy \quad (A.44)$$

for polynomials in $y$, $P^1_{1, n, jk}$ and $P^2_{2, n, jk}$ obtained by our MAPLE-procedure getasintegrand, where $P^1_{1, n, jk}$ is defined as

$$\frac{r_0^2(r_0-1)}{16r_0^4} + \frac{\sqrt{5}r_0^2(r_0-2)^2}{32r_0^4} + \left( \frac{\sqrt{5}(5r_0^2+1)}{8r_0^4} + \frac{\sqrt{5}r_0^2(5r_0^2-2)}{8r_0^4} \right) j + \left( \frac{3r_0^2}{8r_0^4} + \frac{3r_0^2}{8r_0^4} \right) j^2 - \frac{2r_0^2}{4r_0^4} j^3 - \frac{\sqrt{5}r_0^2}{32r_0^4} j^4$$

and $P^2_{2, n, jk}$ as

$$P^2_{2, n, jk}(y) = \frac{2f_0^2(2r_0^2-2y^2)^2+24y^2(2y^2-4r_0^2)}{128y_0^4} + \frac{2f_0^2(2y^2-2r_0^2)^2+24y^2(2r_0^2-4y^2)}{128y_0^4} + \frac{2f_0^2(2y^2-2r_0^2)^2+24y^2(2r_0^2-4y^2)}{128y_0^4} + \frac{2f_0^2(2y^2-2r_0^2)^2+24y^2(2r_0^2-4y^2)}{128y_0^4}$$

By the restriction in $A_{0, n}^\prime$, we obtain that $|y| = O(\sqrt{\log(n)})$, while $\deg(P^1_{1, n, jk}) = 5$ and $\deg(P^2_{2, n, jk}) = 8$. Hence, the integrand is apparently of form $(y + \sqrt{5} f)^2/(4f^2) + O(\sqrt{\log(n)^2/n})$, and thus, eventually in $n$, is maximized—up to $O(\sqrt{\log(n)^2/n})$—for $|y| = k/2$. Even more so, if $f_1 = 0$, the $m^{-1/2}$-term, too, is maximized for $|y| = k/2$. As the highest power in $P_2$, occurring to $y$ in a $j$-term without $f_1$ is 4, the integrand is maximized up to $O(\log(n)^2/n)$ for $|y| = k/2$.

Condition $|y| = k/2$ is equivalent to $j_k(t) \equiv k$ or $j_k(t) \equiv 0$. But this is the case—up to $o(n^{-1})$—if condition (3.5) or (3.6) is in force, as then up to mass of order $o(n^{-1})$ the contamination is either concentrated left
of $-\frac{2k}{n} \sqrt{\log(n)/n}$ or right of $\frac{2k}{n} \sqrt{\log(n)/n}$ for any sample with no more than $k_1 r \sqrt{n}$ contaminations. With respect to Remark A.7, 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 $o(n^{-1})$, $\int_A n^2 g_{k,n,k}(t) \, dt$ is

$$\int_A \left| \frac{1}{\sqrt{n}} \left( f^2 + \frac{f^2}{2} \right) + Q_{1,n,k}(ym^{-1/2} + Q_{2,n,k}(ym^{-1} + \bar{g}(n,k,y)) y \right) dy \quad (A.45)$$

with some skew-symmetric polynomial $\bar{g}$ in $y$ of degree 5 that is uniformly bounded in $n$ on $A_0^1$, 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.

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

$$\int_{A_0^1} n^2 g_{k,n,k}(t) \, dt$$

$$= \frac{15k^2}{16g_0} + \left| \left( \frac{3}{16g_0} + \frac{1}{32g_0} f^2 \right) + \left( \frac{15k_1^2}{128g_0^2} \right) \right| y^{m^{-1/2}} + \left| \left( \frac{3}{32g_0} \right) + \frac{15k_1^2}{128g_0^2} y^{m^{-1}} + \left( \frac{3}{32g_0} \right) \right| y^{m^{-1}}$$

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

$$\sup_n n \text{MSE}[\text{Med}(G^{(0)})] = \frac{1}{4g_0} \left( 1 + r^2 \right) + \frac{1}{\sqrt{2}} \left( 2l_1 r^2 \right) + \frac{\left( 2l_1 r^2 \right)}{\sqrt{n}} + \frac{\left( \log(\log(n)) \right)}{2g_0} + o(n^{-1})$$

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). $\square$

### 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 $M^*_{k,n}$, 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 $\sqrt{n}[\text{Bias}(X_{[m+1:n]}, F^0)] = B_{n,1} + B_{n,2}$ for

$$B_{n,1} = B_{n,1} + B_{n,2}, \quad B_{n,1} = \frac{1}{2\sqrt{n}}, \quad |B_{n,2}| = \frac{1}{8g_0/\sqrt{n}}$$

The same terms but with different signs are obtained for $\sqrt{n}[\text{Bias}(X_{[m+1:n]}, F^0)]$. We only consider $B_{n,1}$ here, which arises no matter if we have symmetry or not and gives the bias corrected version $M^*_{k,n}$ 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

$$M^*_{k,n} := M_n - \frac{1}{\sqrt{n}} B_{n,2} = M_n + \frac{f_1}{8g_0/\sqrt{n}}$$

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 \( \tilde{M}_n \), we need the common law of the pure quantile estimators \( X_{[n]} \) and \( X_{[n+1]} \). So more generally, we start with the common law of \( (Y, Z) := (X_{[r_1]}, X_{[r_2]}) \) for \( 1 \leq r_1 < r_2 \leq n \), \( X_i \overset{i.i.d.}{=} F \), \( i = 1, \ldots, n \) and \( F(dx) = f(x)dx \), see David [1970] pp. 9–10, and in our case \( (m \geq 2m, r_1 \geq 2m, r_2 \geq m + 1) \) leads us to the density of the midpoint estimator \( \tilde{M}_m = (Y + Z)/2 \)

\[ g(t) = (2m)^2 \left( \frac{2m-1}{m} \right) \int_0^\infty \left[ F(2t-u)(1-F(u)) \right]^{m-1} f(u) f(2t-u) du \]  

This gives for \( (2m) \text{MSE}(\tilde{M}_m, F) \), after substituting \( s = 2t - u \), and using Fubini

\[ (2m) \text{MSE}(\tilde{M}_m, F) = 2m^2 \left( \frac{2m-1}{m} \right) \int \int_0^\infty \frac{(u+s)^2}{4} \left[ F(s)(1-F(u)) \right]^{m-1} f(u) f(s) ds du \]

We skip the argument showing how to choose a risk minimizing 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. Analogous arguments as in the preceding cases show that given we have \( k \) observations contaminated to \( \infty \), we get as expression for the (conditional) \( \text{MSE}_{G_k, \delta} \):

\[ (2m) \text{MSE}_{G_k, \delta} = (2m)(2m-k) \int_{m-k}^{m-1} F(u)^{m-1} \left( 1 - F(u) \right)^{m-1} f(u) \times \]

\[ \times \int_0^\infty \left( \frac{m-k-1}{m} \right) f(u)^{(m-k)(u+s)^2} 4F(u) (1 - F(u))^{(m-k)(u+s)^2} f(s) ds du \]

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

\[ A(s, u) := \left( F(u) - F(s) \right) / F(u) \]

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

\[ \delta(a) := \sup \{ s < u | F(s) \leq (1 - \alpha) F(u) \} \]

in fact we may restrict the inner integral to

\[ H_k(a) = o(m^{-1}) + \frac{m-k}{4F(U)} \int_{\alpha(a)}^U (u+s)^2 (1-A(s,u))^{(m-k)(u+s)^2} f(s) ds \]

But then expanding \( \log(1-A(s,u)) \), and in order to get the right order for the expansion substituting \( u = \bar{u} / \sqrt{m}, s = \tilde{s} / \sqrt{m} \)—according to case (I), i.e., \( |u| \leq \sqrt{\log(m)/m} \). Thus, for polynomials \( Q_r \) in \( \bar{s}, \tilde{s} \) defined in analogy to \( Q_k \) in the to odd-sample case and with may be looked up in the MAPLE script,

\[ A(s, u) = 2 \frac{h_{(m-k)}}{\sqrt{m}} + \frac{h_{(m-k)}}{\sqrt{m}} + \frac{Q_{(m-k)}}{\sqrt{m}} + \frac{Q_{(m-k)}}{\sqrt{m}} + O\left( \frac{\log(m)}{m} \right)^2 \]

Hence we get

\[ (m-k-1) \log(1-A(s,u)) - \sqrt{m} \log(1-A(s,u)) = \log_{H21}(s, u) + o(\sqrt{\log(m)/m}) \]

for some function \( \log_{H21} \), the exact expression of which may be produced in the corresponding MAPLE script. Thus, denoting the term \( \exp(2 \sqrt{m} \log_2 (\bar{s} - \tilde{s})) \) by \( H_2(s, u) \), we get

\[ (1 - A(s,u))^{(m-k)(u+s)^2} = H_2(s, u) \exp(\log_{H21}) \times (1 + o(\sqrt{\log(m)/m})) \]

Now, if we write \( H_{2,2}(s,u) \) for \( s + u^2 f(s) \), and \( H_{2,2}(s,u) \) for \( \exp(\log_{H21}) \), and if we introduce \( H_{2,2}(s,u) := H_{2,2}(s,u)H_{2,2}(s,u) \), we get

\[ 4F(u) H_k(u) = o(n^{-2}) + \int_{\alpha(a)}^U H_2(s, u) H_{2,2}(s, u) ds \]
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

$$4F(u) H_k(u) = o(n^{-1}) + \frac{2}{\sqrt{m}} \sum_{i=0}^{\frac{n}{2}} \left[ 1 - \frac{1}{2} \int (2 \sqrt{m})^{i+1} H_{2i}(x,u) \frac{\partial^i}{\partial x^i} H_{2i}(x,u) \right]_0^n$$  \hspace{1cm} \text{(A.55)}$$

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

$$H_k(t) = t^2 + \frac{1}{2} ((\frac{1}{3^2} - 1)t^2 - \frac{1}{2^2} t - \frac{1}{3^2} + \frac{1}{2} \frac{1}{3^2}) + o(n^{-2})$$

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

$$\sup_{G^m} n \text{MSE}(\hat{\mu}_n, G^m) = n \sum_{k=0}^{m} \int H_k(t) \, g_{n,k}(t) \, dt \, P(K = k) + o(n^{-1})$$  \hspace{1cm} \text{(A.56)}$$

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

\textbf{A.9 Proof of Proposition 3.9}

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

$$A_{k,j} := \left\{ \sum_{i} U_i(2 I(X_i \leq t) - 1) \leq k - 1 \right\}$$   \hspace{1cm} \text{(A.57)}$$

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

$$P(A_{k,j} \, | \, K = k) \geq p_0$$   \hspace{1cm} \text{(A.58)}$$

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

$$n \left\{ \text{MSE}(\hat{\mu}_n, G^m_n) \mid K = k \right\} - \text{MSE}(\hat{\mu}_n, G^m_0) \mid K = k \right\} \geq \frac{p_0}{2 f_0} \int_{\sqrt{\log n}}^{k_2 \sqrt{\log n}} \left( s/ \sqrt{n} + k/ \sqrt{2} - 1/(2 f_0) \right) \varphi(s) \, ds + o(n^{-1})$$

But for $0 < a_1 < a_2 < \infty$, $\varphi(a_1)/a_2 - \varphi(a_2)/a_2 \leq \int_{a_1}^{a_2} \varphi(s) \, ds$, so that with $a_1 = 2 \sqrt{\log n}$, $a_2 = k_2 \sqrt{\log n}$, and as $\varphi(a_2) = o(n^{-1})$,

$$n \left\{ \text{MSE}(\hat{\mu}_n, G^m_n) \mid K = k \right\} - \text{MSE}(\hat{\mu}_n, G^m_0) \mid K = k \right\} \geq \frac{p_0}{2 \sqrt{2} \pi \, f_0} + o(n^{-1})$$

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. \hfill \Box

\textbf{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 Jureˇckov´a 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.

Jureˇckov´a 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/