# Higher Order Asymptotics on Shrinking Neighborhoods

Peter Ruckdeschel

UNIVERSITÄT BAYREUTH
Mathematisches Institut
Peter.Ruckdeschel@uni-bayreuth.de
www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL

Seminar at EPFL
October 27, 2006

## Outline of Sections I – III



## Outline of Sections IV and V



## Outline

## Ideal Setup

Setup: inference on parameter $\theta$ in a model for i.i.d. observations

$$\mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\} \qquad \Theta \subset \mathbb{R}^k, \qquad \mathcal{P} \text{ “smooth”}$$

- common robust technique:
  use first order *von-Mises (vM) expansion*

Definition

**influence curves at $P_\theta$:**

$$\Psi_2(\theta) = \{\psi_\theta \in L_2^k(P_\theta) \,|\, \mathrm{E}_\theta\, \psi_\theta = 0,\ \mathrm{E}_\theta\, \psi_\theta \Lambda_\theta^\tau = \mathbb{I}_k\}$$

**asymptotically linear estimators:**

$$\sqrt{n}\,(S_n - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\theta(x_i) + o_{P_\theta^n}(n^0)$$

## Infinitesimal Robust Setup

**Shrinking neighborhoods (Rieder[81,94], Bickel[83])**

$$U_c(\theta, r, n) = \big\{(1 - r/\sqrt{n}\,)_+ P_\theta + (1 \wedge r/\sqrt{n}\,)\, R \,\big|\, R \in \mathcal{M}_1(\mathcal{A})\big\}$$

**Robust optimality problem:** $\sup_{Q \in U_c} \mathrm{MSE}_Q(\psi_\theta) = \min!$

here: $\sup_{Q \in U_c} \mathrm{MSE}_Q(\psi_\theta) = \mathrm{E}_\theta\, |\psi_\theta|^2 + r^2 \sup |\psi_\theta|^2$

**Thm.s 5.5.1 and 5.5.7 (b), Rieder[94]**

*unique solution is an IC $\tilde\eta_\theta$ of Hampel-type, i.e.;*

$$\tilde\eta_\theta = (A_\theta \Lambda_\theta - a_\theta) w \qquad w = \min\big\{1, b_\theta / |A_\theta \Lambda_\theta - a_\theta|\big\}$$

*with $A_\theta$, $a_\theta$, $b_\theta$ such that* $\mathrm{E}_\theta\, \tilde\eta_\theta = 0$, $\mathrm{E}_\theta\, \tilde\eta_\theta \Lambda_\theta^\tau = \mathbb{I}_k$, *and*

$$(\mathrm{MSE}) \qquad r^2 b_\theta = \mathrm{E}_\theta\, \big(|A_\theta \Lambda_\theta - a_\theta| - b_\theta\big)_+$$

## Limitations of First Order Approach

- So far: asymptotics is of first order, for both ALE and MSE
- Limitations
  - (Technicality: in order to force convergence of the risk: modification of the loss function by clipping)
  - no indication for the quality/speed of the convergence — to what degree do radius $r$, sample size $n$ and clipping height $b$ affect the approximation?
  - no indication which construction achieving an optimally–robust IC asymptotically to take

## Outline

## A "Historic" Aside I. . . : Lausanne 2003

- simulational evidence for reasonable results of the infinitesimal setup was reported:
  - good convergence of the MSE's to the asymptotic ones for the ideal model (tentative rate $1/n$)
  - (slow) convergence of the MSE's to the asymptotic ones for the contaminated model (tentative rate $1/\sqrt{n}$)
  - convergence speed gets slower for $r$, $k$ increasing
  - good convergence of the relative MSE's to the asymptotic ones for the contaminated model (tentative rate $1/n$ or faster)
  - order of the relative asymptotic MSE's for different $k$ is preserved largely (at most another $k$ yields a performance gain of 3% w.r.t. $k_0$ — for $n = 5$ and $r = 0.1$)
  - never large differences between constructions, for the best $r$ of order $10^{-2}$ or smaller

## A "Historic" Aside II. . . : Lausanne 2003

- preliminary convergence results were presented:

Theorem (for both $n$ odd and even — R.[05(a)])

*for the sample median in the ideal Model*

$$n\,\mathrm{MSE}(\mathrm{Med}_n, \mathcal{N}(0,1)) = \tfrac{\pi}{2}\Big[1+\big(\tfrac{\pi}{2}-2\underbrace{[-1]}_{\text{for midpoint if } n \text{ even}}\big)/n\Big]+\mathrm{o}(\tfrac{1}{n})$$

*uniform conv. for* $\mathrm{Med}_n$ *on Nbd about the ideal Model* $\mathcal{N}(0,1)$

$$n\,\sup_{F^{\mathrm{real}}} \mathrm{MSE}(\mathrm{Med}_n, F^{\mathrm{real}}) = \tfrac{\pi}{2}(1+r^2)(1+\frac{2r}{\sqrt{n}}+\mathrm{o}(\tfrac{1}{\sqrt{n}}))$$

## Different Constructions With Same IC

- By means of first order asy. no distinction possible between
  - M-estimator (does not dependent on $\theta_n^{(0)}$!):

$$\theta_n^{(z)} \quad \text{s.t.} \quad g_n(\theta_n^{(z)}) = 0 \quad \text{for } g_n(\theta) = \sum_{i=1}^{n} \eta_\theta(X_i),$$

  - $k$-step-estimator: to some starting estimator $\theta_n^{(0)}$,

$$\theta_n^{(k)} := \theta_n^{(k-1)} + \frac{1}{n}\sum_{i=1}^{n} \eta_{\theta_n^{(k-1)}}(X_i)$$

⇝ central question of this talk:

**Which one—$k$-step- or M-estimator—has smaller risk for fixed $n$?**

## Existing Approaches To Assess This Question

- *vM-expansion* (Jurečkova and varying coauthors, [83–97])

  idea: for two estimators $S_n, S_n'$, expand $\Delta_n = S_n - S_n'$ to higher order (for smooth ICs)

  but need not exist (e.g. median); then: *Bahadur-Kiefer representation* for the remainder

  – due to correlation: $\mathcal{L}(\Delta_n)$ of little help for comparison of $\mathcal{L}(S_n)$, $\mathcal{L}(S_n')$
- *distributional expansion (Edgeworth / Saddlepoint approx.)* (e.g. Ronchetti and Welsh [02])
  - more flexible but (Saddlepoint approx.) less explicit analytically
  - \+ suffices for (MSE-)risk under uniform integrability

up to now: no uniform statements on neighborhoods

## Uniform Expansions of the MSE I

**Theorem (R. [05(a,b,c)])**

Let *$\theta \mapsto \eta_\theta$ be smooth in $L_1(P_\theta)$,*
*$S_n$ be an M- or a k-step-estimator to $\eta_\theta$, and*

let *starting estim. $\theta_n^{(0)}$ for the k-step-estimator be*

- *uniformly $n^{1/4+\delta}$-consistent on $\tilde U_c$ for some $\delta > 0$*
- *uniformly square-integrable in n and on $\tilde U_c$*

Then

$$\mathrm{maxMSE}(S_n) \;:=\; n \sup_{Q_n \in \tilde U_c(r)} \mathrm{MSE}(S_n)$$
$$= \boxed{A_0 + \tfrac{r}{\sqrt{n}} A_1 + \tfrac{1}{n} A_2 + \mathrm{o}(\tfrac{1}{n})}$$

*for $A_0 = \mathrm{E}_\theta |\eta_\theta|^2 + r^2 \sup |\eta_\theta|^2$ and $A_1$, $A_2$ are constants depending on $\eta_\theta$, $r$, and, for k-step-est., also on $\theta_n^{(0)}$*

## As to Uniform Integrability:

**Breakdown-restricted samples**

- by breakdown-point type argument: no uniform convergence of MSE on neighborhoods $U_c(\theta, r, n)$ for $r > 0$

$\rightsquigarrow$ sample-wise restriction of the neighborhoods, *conditioned on # contaminations in sample* $\rightsquigarrow \tilde U_c(\theta, r, n)$:

s.t. *percentage of contaminations in such samples smaller than the finite-sample breakdown-point of most robust estimator $S_n^\flat$.*

e.g. in the location case, samples with more than 50% contaminiations are excluded

- by *Hoeffding:* restriction is asymptotically exponentially negligible

## Uniform Expansions of the MSE II

**Exact expressions for $A_1$ for 1-step-estimator in one dimension**

Let $\eta_\theta$ bounded and two times differentiable in $L_1(P_\theta)$,
$\theta_n^{(0)} = \theta + \frac{1}{n} \sum \tilde\eta_\theta(x_i) + \mathrm{o}_{L_1(\tilde U_c)}(n^{-1/2})$ for a bounded IC $\tilde\eta_\theta$,

Then

$$\begin{aligned} A_1 \;=\; & 2\,\mathrm{Cov}_\theta(\eta_\theta, \tilde\eta_\theta) - \mathrm{Var}_\theta\, \eta_\theta^2 + b_\theta^2 \\ & + 2 b_\theta^2 \tfrac{d}{dt} \mathrm{Cov}_\theta(\eta_t, \tilde\eta_\theta)\big|_{t=\theta} + 2\tilde b_\theta^2 \tfrac{d}{dt} \mathrm{Var}_\theta\, \eta_t\big|_{t=\theta} \\ & + \tfrac{d^2}{dt^2} \mathrm{E}_\theta\, \eta_t\big|_{t=\theta} \Big[ b_\theta \mathrm{Var}_\theta\, \tilde\eta_\theta + 2\tilde b_\theta \mathrm{Cov}_\theta(\eta_\theta, \tilde\eta_\theta) \Big] \\ & + r^2 \tilde b_\theta b_\theta \; \big[2 + \tilde b_\theta \tfrac{d^2}{dt^2} \mathrm{E}_\theta\, \eta_t\big|_{t=\theta}\big] \end{aligned}$$

where $\quad b_\theta = \sup|\eta_\theta|, \quad \tilde b_\theta = \limsup_{\varepsilon \downarrow 0} \sup |\tilde\eta_\theta|\, \mathrm{I}(|\eta_\theta| \geq b_\theta - \varepsilon)$

M-est put $\tilde\eta_\theta = \eta_\theta$

## Outline

## Uniform MSE-Expansions for M-estimators and Median

**Theorem (R. [05(a,b)])**

Let *$\eta_\theta$ be a bounded, monotone IC s.t. $\theta \mapsto \mathrm{E}_\vartheta\, \psi_\theta^j$ smooth*
*$P_\theta$ have at least polynomial tails*
*$S_n$ be an M-estimator to $\eta_\theta$*

Then
$$\mathrm{maxMSE}(S_n) := n \sup_{Q_n \in \tilde{U}_c(r)} \mathrm{MSE}(S_n) = A_0 + \frac{r}{\sqrt{n}} A_1 + \frac{1}{n} A_2 + \mathrm{o}(\tfrac{1}{n})$$

*for $A_0 = \mathrm{E}_\theta\, |\eta_\theta|^2 + r^2 \sup |\eta_\theta|^2$ and $A_1$, $A_2$ are constants depending on $\eta_\theta$ and $r$*

## Comments

**Theorem (R. 2005(a)/2005(b))**

*In the last theorem, maximal MSE is already attained if contam. is concentrated strictly right [left] of $\eta_\theta^{-1}(\sup|\eta_\theta|) \pm b\sqrt{2\log(n)/n}$.*

- special proof for median due to violation of Cramér-condition
- also possible:
  - treatment of bias and variance, separately
  - over-/under-shoot probability-risk
- cross-checks: simulations, numerical evaluations
- compared to saddlepoint-approximations, c.f. Field and Ronchetti [90] much more explicit terms
- gives justification (ex post) for good approximation quality of Fraiman et al [01]

## Sketch of Proof I

- use $Q_n(S_n \geq t) = Q_n(\sum_i \psi(X_i - t) > 0)$ essentially
- conditioning w.r.t. the number $K$ of contaminated observations
- cond. w.r.t. the actual contam. $\tilde{T}_{m,k} = \sum_{U_i=1} \psi(X_i^{\mathrm{di}})$
- partitioning the integrand of the conditional MSE,

<table>
<tr><th></th><th>$K < k_1 r\sqrt{n}$</th><th>$k_1 r\sqrt{n} \leq K < n/2$</th><th>$K \geq n/2$</th></tr>
<tr><td>$|t| \leq k_2 b^2 \log(n)/n$</td><td>(I)</td><td rowspan="2">(II)</td><td rowspan="3">excluded</td></tr>
<tr><td>$k_2 b^2 \log(n)/n < |t| \leq C n^{k_1'}$</td><td>(III)</td></tr>
<tr><td>$|t| > C n^{k_1'}$</td><td colspan="2">(IV)</td></tr>
</table>

for any fixed $k_1, k_1' > 1$, $k_2 > 2$

- showing negligibility of cases (II),(III), and (IV)
- using an Edgeworth expansion on (I)

## Sketch of proof II

- change of variables $t = t(s)$ to extract argument $s$ from exp —(implicit function theorem!)
- with `MAPLE`: collecting terms
- identification of the least favorable contamination
- integration of $s$ — conditional on $K = k$
- integration of $K$

## Consequence: Second Order Optimality I

**Corollary**

Let *$F$ and $\psi$ be symmetric:*

Then $A_1 = 2r^2b^2 + v_0^2 + b^2$ *and maximal risk is*
$R_1(S_n) = r^2b^2 + {v_0}^2 + \frac{r}{\sqrt{n}} A_1$

Consequence:
Second order optimal (s-o-o) IC is of Hampel form

$$A\Lambda \min\{1, c_1/|\Lambda|\}$$

with s-o-o clipping height $c_1$ determined as

$$r^2 c_1 \left(1 + \frac{r^2+1}{r^2 + r\sqrt{n}}\right) = \mathrm{E}(|\Lambda| - c_1)_+$$

## Consequence: second order optimality II

If $h(c) := \mathrm{E}(|\Lambda| - c)_+$ is differentiable in the f-o-o $c_0$,

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

$\Longrightarrow$ As $h' < 0$, $c_1 < c_0$ always

i.e.; *first order asymptotics is too optimistic*

- as $c_1$ is optimal, $R_1$ behaves locally as a parabola with vertex in $c_1$; hence the risk-improvement of $c_1$ compared to $c_0$ is $\mathrm{O}(1/n)$
- same goes for t-o-o clipping height $c_2 \Longrightarrow$ risk-improvement of $c_2$ compared to $c_1$ is $\mathrm{O}(1/n^2)$

## Optimal $c$'s and corresp. (numerically) exact maxMSE I

| $r$ | | $n=5$ | $n=10$ | $n=30$ | $n=100$ |
|---|---|---|---|---|---|
| 0.1 | actual rad. | 0.04 | 0.03 | 0.02 | 0.01 |
| | $c_0$ | 1.948 | 1.948 | 1.948 | 1.948 |
| | $\mathrm{relMSE}_n^{\mathrm{ex}}(c_0)$ | 8.679% | 4.065% | 1.340% | 0.448% |
| | $c_1$ | 1.394 | 1.484 | 1.611 | 1.724 |
| | $\mathrm{relMSE}_n^{\mathrm{ex}}(c_1)$ | 0.833% | 0.207% | 0.027% | 0.010% |
| | $c_2$ | 1.309 | 1.428 | 1.585 | 1.713 |
| | $\mathrm{relMSE}_n^{\mathrm{ex}}(c_2)$ | 0.332% | 0.066% | 0.008% | 0.006% |
| | $c_{\mathrm{FZY}}$ | 1.368 | 1.370 | 1.610 | 1.756 |
| | $\mathrm{relMSE}_n^{\mathrm{ex}}(c_{\mathrm{FZY}})$ | 0.658% | 0.002% | 0.026% | 0.031% |
| | $c_{\mathrm{ex}}$ | 1.167 | 1.358 | 1.560 | 1.704 |
| | $\mathrm{MSE}_n(c_{\mathrm{ex}})$ | 1.388 | 1.239 | 1.151 | 1.107 |

$c_0$ f-o-o: by equation we just saw
$c_1$ s-o-o: by equation we just saw
$c_2$ third order: num. optimization of MSE among Hampel-type ICs
$c_{\mathrm{FZY}}$ num. optimization of a proposal by Fraiman et al.
$c_{\mathrm{ex}}$ num. optimization of the (num.) exact MSE

## Optimal $c$'s and corresp. (numerically) exact maxMSE II

| $r$ | | $n=5$ | $n=10$ | $n=30$ | $n=100$ |
|---|---|---|---|---|---|
| 1.0 | actual rad. | 0.45 | 0.32 | 0.18 | 0.10 |
| | $c_0$ | 0.436 | 0.436 | 0.436 | 0.436 |
| | $\mathrm{relMSE}_n^{\mathrm{ex}}(c_0)$ | 2.716% | 3.132% | 0.746% | 0.149% |
| | $c_1$ | 0.320 | 0.340 | 0.369 | 0.394 |
| | $\mathrm{relMSE}_n^{\mathrm{ex}}(c_1)$ | 1.411% | 1.610% | 0.251% | 0.021% |
| | $c_2$ | 0.255 | 0.291 | 0.342 | 0.382 |
| | $\mathrm{relMSE}_n^{\mathrm{ex}}(c_2)$ | 0.876% | 0.999% | 0.123% | 0.006% |
| | $c_{\mathrm{FZY}}$ | – | 0.281 | 0.344 | 0.387 |
| | $\mathrm{relMSE}_n^{\mathrm{ex}}(c_{\mathrm{FZY}})$ | – | 0.892% | 0.132% | 0.012% |
| | $c_{\mathrm{ex}}$ | 0.001 | 0.125 | 0.286 | 0.366 |
| | $\mathrm{MSE}_n(c_{\mathrm{ex}})$ | 12.627 | 8.445 | 4.948 | 3.787 |

$c_0$ f-o-o: by equation we just saw
$c_1$ s-o-o: by equation we just saw
$c_2$ third order: num. optimization of MSE among Hampel-type ICs
$c_{\mathrm{FZY}}$ num. optimization of a proposal by Fraiman et al.
$c_{\mathrm{ex}}$ num. optimization of the (num.) exact MSE

## Outline

New Concepts in Robust Statistics

Minimax-Radius

Second Order Minimax-Radius

Cniper Contamination

Cniper Contamination and Second Order Asymptotics

## Unknown Radius $r$: Minimax-Radius

- situation: $r$ not known, only available information $r \in [r_l, r_u]$
- relative inefficiency of $\eta_r$ when used at radius $s$:

$$\rho(r,s) := \max_{\rm nbd} \operatorname{asRisk}(\eta_r, s) / \max_{\rm nbd} \operatorname{asRisk}(\eta_s, s)$$

- minimax radius/inefficiency:
  $r = r_0$ such that $\hat\rho(r)$ is minimal for $\hat\rho(r) := \sup_{s\in[r_l,r_u]} \rho(r,s)$

**Theorem (Radius-minimax procedure [R.:Ri:04])**

*Assume that maximal asymptotic risk on nbd is representable as*
$\tilde G(\eta, r) = G(r\omega_\eta, \sigma_\eta)$ *for*

- $\sigma_\eta^2 = \mathrm{E}_P\,|\eta|^2$, $\omega_\eta = \sup_{Q \textit{ in nbd}} |\,\mathrm{E}_Q\,\eta|$
- $G = G(w,s)$ *convex, isotone in both arguments*
- $G$ *homogeneous,i.e.;* $G(\nu w, \nu s) = \nu^\alpha G(w,s)$

*For all such* $G$*, the radius-minimax IC does* **not** *depend on* $G$*!*

## Second Order Minimax-Radius

- Set $R_1(\psi, r, n) := r^2 \sup|\psi|^2 + \mathrm{E}\,\psi^2 + \frac{r}{\sqrt{n}} A_1$ and let $c_1(r,n)$ the s-o-o $c$; then determine the *s-o-minimax-radius* $r_1 = r_1(n)$ as minimizer of

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

- Illustration at Gaussian location model for $r_l = 0$, $r_u = \infty$

| | $n=5$ | $n=10$ | $n=100$ | $n=\infty$ |
|---|---|---|---|---|
| $r_1$ | 0.390 | 0.449 | 0.559 | 0.621 |
| $c_1(r_1)$ | 0.776 | 0.749 | 0.722 | 0.718 |
| $\rho_1(r_1)$ | 16.27% | 17.08% | 17.96% | 18.07% |

- **So if $r$ is completely unknown, use the M-estimator to $\eta_c$ for $c \approx 0.7$ — you will never have a larger inefficiency than the limiting** 18%**!**

## Cniper Contamination

- Huber [97], p. 62, complains "...the considerable confusion between the respective roles of diagnostics and robustness. The purpose of robustness is to safeguard against deviations from the assumptions that are near or below the limits of detectability."
- In R. [06]: determination of these limits in a statistical way, using binomial maximin tests, giving exact critical rate $1/\sqrt{n}$
- Idea: Among risk-maximizing contamination(s) determine the "*most innocent appearing least favorable contamination*"

⇝ H. Rieder: cniper–contamination: Being of Ianus-type, it pretends to be *nice* but in fact is already *pernicious*.

## Cniper Contamination and Second Order Asymptotics

### Proposition

Let $Q_n(x) := (1 - r/\sqrt{n})F + r/\sqrt{n}\,\mathrm{I}_{\{x\}}$
*and* $\mathrm{asMSE}_1(S, Q)$ *the s-o as. MSE of* $S$ *under* $Q$.
*Define* $x_1 = x_1(n)$ *as the minimal* $x > 0$ *such that*

$$\mathrm{asMSE}_1(S_n^{(c_1)}, Q_n(x)) = \mathrm{asMSE}_1(\hat{S}_n, Q_n(x))$$

*for* $S_n^{(c_1)}$ *the s-o-o M-estimator and* $\hat{S}_n$ *is the MLE.*

Then *one can show:* $(S_n^{(c_1)}, Q_n(x_1(n)))$ *is a saddlepoint.*

**Illustration: one-dim. Gaussian location (known scale)**

| $n$ | 5 | 10 | 100 | $\infty$ |
|---|---|---|---|---|
| $r_1(n)$ | 0.390 | 0.449 | 0.559 | 0.621 |
| $c_1(r_1, n)$ | 0.776 | 0.749 | 0.722 | 0.718 |
| $x_1(n)$ | 2.937 | 2.465 | 1.800 | 1.524 |

## Outline



## Specialization: One-dim. Symmetric Location

### Proposition

Let $\Lambda_\theta(-\,\cdot\,) = -\Lambda_\theta(\,\cdot\,)$

- $\tilde{\eta}_\theta$ *MSE-optimal IC to radius* $r$ *(with clipping height* $\tilde{b}_\theta$*)*
- $\eta_\theta^{(b_\theta)} = A_\theta\Lambda_\theta \min\{1, \frac{b_\theta}{|A_\theta\Lambda_\theta|}\}$ *for some* $0 < b_\theta < \tilde{b}_\theta$.
- $S_n, S_n'$ *be the resp. M- and 1-step-estimator to* $\tilde{\eta}_\theta$, *with* $\theta_n^{(0)}$ *an ALE with IC* $\eta_\theta^{(b_\theta)}$

Then $\quad \mathrm{maxMSE}(S_n') = \mathrm{maxMSE}(S_n) + \mathrm{o}(n^{-1/2})$

### Remark

*No general statement to our central question:*
*If IC is of Hampel-type and first order MSE-suboptimal, then both situations* $\mathrm{maxMSE}(S_n') \lesseqgtr \mathrm{maxMSE}(S_n) + \mathrm{o}(n^{-1/2})$ *may occur.*

## Higher Order Comparison of maxMSE

Uniform expansion of MSE allows the following comparison

### Theorem (R. 2005(b))

Let $\theta \mapsto \eta_\theta$ *be* $k$ *times differentiable in* $L_1(P_\theta)$.
$S_n, S_n'$ *be the resp. M- and* $k$*-step estimator to* $\eta_\theta$.
$\theta_n^{(0)}$ *to* $S_n'$ *be uniformly consistent and integrable as before*

Then *there exist expansions of order* $k$ *of* $\mathrm{maxMSE}$ *for* $S_n, S_n'$ *and*

$$\mathrm{maxMSE}(S_n') = \mathrm{maxMSE}(S_n) + \mathrm{o}(n^{-(k-1)/2})$$

- preceding theorem covers $n^{1/3}$-consistent $\theta_n^{(0)}$s like *Least-Median-of-Squares*-regression estimators
- we apply theorem to $k = 3$, as explicit expressions for expansions available only up to order 3
- extension to non-$L_1$-smooth ICs like Hampel-type-ICs for $k = 3$ by ad-hoc methods

## Optimal Robustness Combined With High Breakdown

- use of high-breakdown estimators *slower* than $n^{-(1/4+\delta)}$

### Proposition (R.05: Acceleration of slow starting estimators)

Let $\tilde\theta_n^{(0)}$
- *uniformly $n^\alpha$-consistent on $\tilde U_c(r)$ for some $0<\alpha\le 1/4$*
- *uniformly square-integrable as in the theorem*

Then *an $m=\lceil -1-\log_2\alpha\rceil$-step-estimator $\tilde\theta_n^{(m)}$ to* any *$L_1(P_\theta)$-smooth IC with $\theta_n^{(0)}=\tilde\theta_n^{(0)}$ is uniformly integrable and*

*becomes $n^{1/4+\delta}$-consistent,*

$\Longrightarrow$ *is admitted as starting estimator in preceding theorem*

- high breakdown of $\tilde\theta_n^{(0)}$ is inherited to $k$-step-estimators (not true for M-estimators!)

$\Longrightarrow$ optimal uniform efficiency + optimal breakdown point

## Empirical Results: Simulation Design

- ideal model: $\mathcal P=\mathcal N(\theta,1)$ at $\theta=0$
- $M=10000$ runs; sample sizes: $n=5,10,30,50,100$
- contamination radii: $r=0.1,0.25,0.5,1.0$
- contaminating distribution: Dirac at point 100
- ICs: Huber-type to $c=0.5,0.7,1,1.5,2$
- estimators:
  - M-estimator and
  - 1-Step-estimator with sample median as starting estimator

## Empirical Results: Simulation Results I

### Empirical and asymptotic $\mathrm{maxMSE}$ at $n=30$, $c=0.5$

| $r$ / $r/\sqrt{n}$ | M/1step | simulation $\overline{\mathrm{maxMSE}}_n$ | [low; up] | asymptotics $n^0$ | $n^{-1/2}$ | $n^{-1}$ |
|---|---|---|---|---|---|---|
| 0.00 | 1step | 1.270 | [1.235 ;1.306] | 1.263 | 1.263 | 1.258 |
| 0.00 | M | 1.272 | [1.237 ;1.307] | 1.263 | 1.263 | 1.259 |
| 0.25 | 1step | 1.553 | [1.510 ;1.596] | 1.369 | 1.519 | 1.544 |
| 0.05 | M | 1.545 | [1.502 ;1.588] | 1.369 | 1.514 | 1.532 |
| 1.00 | 1step | 5.357 | [5.214 ;5.500] | 2.967 | 4.127 | 4.772 |
| 0.18 | M | 5.362 | [5.219 ;5.505] | 2.967 | 4.132 | 4.652 |

$\overline{\mathrm{maxMSE}}_n$: average of emp. risks, low/up: emp. 95% confidence interval
asymptotics taken from leading terms of the preceding expansions:
$A_0\,[\,+\,rn^{-1/2}\,A_1\,(\,+\,n^{-1}A_2)]$, respectively

## Empirical Results: Simulation Results II

### Number of iterations $I_n$ needed for M-Estimator at $n=30$ and $c=0.5$, as well as $n=50$ and $c=2.0$

| $r$ | $n=30$ and $c=0.5$: $\bar I_n$ | [low; | up] | $n=50$ and $c=2.0$: $\bar I_n$ | [low; | up] |
|---|---|---|---|---|---|---|
| 0.00 | 7.00 | [ 5; | 9] | 5.56 | [ 4; | 7] |
| 0.10 | 8.62 | [ 5; | 12] | 7.17 | [ 4; | 10] |
| 0.25 | 9.93 | [ 5; | 12] | 8.54 | [ 5; | 10] |
| 0.50 | 10.56 | [ 7; | 12] | 9.36 | [ 6; | 10] |
| 1.00 | 10.70 | [ 8; | 13] | 9.74 | [ 8; | 11] |

$\Longrightarrow$ *statist. unjustified computation time compared to 1-step*

# Bibliography I


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

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

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

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

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

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

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

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

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

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

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

[12] Huber P.J. (1968): Robust confidence limits. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **10**: 269–278.

[13] —— (1981): *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley.

[14] —— (1997): *Robust statistical procedures.*, Vol. 68 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2. edition.


# Bibliography II


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

[16] Ibragimov I. and Linnik Y. (1971): *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing Company, Groningen. Edited by J.F.C. Kingman.

[17] Jurečková J. (1983): Robust estimators of location and regression parameters and their second order asymptotic relations. In: *Information theory, statistical decision functions, random processes, Trans. 9th Prague Conf., Prague 1982* Vol. A, p. 19–32.

[18] Jurečková, J. and Sen P.K. (1997): Asymptotic representations and interrelations of robust estimators and their applications. In: *Robust inference.* (Maddala, G. S. et al. (Eds.)), Vol. 15 of *Handbook of Statistics* p. 467–512.

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

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

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

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

[23] —— (1985): *Asymptotic expansions for general statistical models. With the assist. of W. Wefelmeyer*, Vol. 31 of *Lecture Notes in Statistics*. Springer-Verlag.

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

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

[26] Rieder H. (1980): Estimates derived from robust tests. *Ann. Stat.*, **8**: 106–115.

[27] —— (1994): *Robust asymptotic statistics*. Springer Series in Statistics. Springer.


# Bibliography III


[28] Rieder H., Kohl M. and Ruckdeschel P. (2001): The Costs of not Knowing the Radius. To appear in Statistical Methods and Applications; also available in `http://www.uni-bayreuth.de/departments/math/org/mathe7/RIEDER/pubs/RR.pdf`.

[29] Ronchetti E. and Welsh A.H. (2002): A journey in single steps: robust one-step M-estimation in linear regression *Journal of Statistical Planning and Inference*, **103**(1): 287–310.

[30] Rousseeuw P.J. (1984): Least median of squares regression. *J. Am. Stat. Assoc.*, **79**: 871–880.

[31] Ruckdeschel P. (2005a): Higher order asymptotics for the MSE of the median on shrinking neighborhoods. Unpublished manuscript. Also available in `http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/medmse.pdf`.

[32] —— (2005b): Higher order asymptotics for the MSE of M-estimators on shrinking neighborhoods. Unpublished manuscript. Also available in `http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/mest1.pdf`.

[33] —— (2005c): Higher order asymptotics for the MSE of *k*-step estimators on shrinking neighborhoods. Unpublished manuscript.

[34] —— (2006): A Motivation for $1/\sqrt{n}$–Shrinking–Neighborhoods. *Metrika* **63**(3) 295–307. Also available in `http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/whysqn.pdf`.

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

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

[37] Ruckdeschel P., Kohl M., Stabla T. and Camphausen F. (2006): `S4` Classes for Distributions. *R News* **6**(2) 2–6. Updated and extended version in `http://www.uni-bayreuth.de/departments/math/org/mathe7/DISTR`.

[38] Yohai V.J. (1987): High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.*, **15**(2): 642–656.