k-step- and M-estimators – a comparison of MSE by uniform higher order asymptotics

Peter Ruckdeschel

UNIVERSITÄT BAYREUTH

Fakultät für Mathematik und Physik

Peter.Ruckdeschel@uni-bayreuth.de

www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL

ICORS06, Lisbon

July, 18 2006

Ideal Setup

Setup: inference on parameter θ 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_θ:

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

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 Hampel-type IC $\tilde\eta_\theta$, 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_θ, a_θ, b_θ such that $\mathrm{E}_\theta\, \tilde\eta_\theta = 0$, $\mathrm{E}_\theta\, \tilde\eta_\theta \Lambda_\theta^\tau = \mathbb{I}_k$, *and*

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

Different constructions with same IC

- So far: asymptotics is of first-order, for both ALE and MSE

$\rightsquigarrow$ no distinction possible between

  - M-estimator (does not dependent on $\theta_n^{(0)}$!):

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

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

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

$\rightsquigarrow$ 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. 2005(a)/2005(b))

Let $\theta \mapsto \eta_\theta$ *be smooth in* $L_1(P_\theta)$,

S_n *be an M- or a k-step-estimator to* η_θ*, 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

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

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

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

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

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

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

Remark

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

$\theta_n^{(0)}$ *to* S_n' *be uniformly consistent and integrable as before*

Then *there exist expansions of order* k *of* 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^{α}*-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* <u>*any*</u> $L_1(P_\theta)$*-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

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

Simulation Results I

Empirical and asymptotic 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

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	Iterations					
	$n = 30$ and $c = 0.5$			$n = 50$ and $c = 2.0$		
	$\bar{I}_n$	[low;	up]	$\bar{I}_n$	[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

H. Rieder (1994):
Robust Asymptotic Statistics. Springer.

P. Ruckdeschel (2005(a)):
Higher order asymptotics for the MSE of M estimators on shrinking neighborhoods. In preparation. Also available in http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/mest1.pdf.

P. Ruckdeschel (2005(b)):
Higher order asymptotics for the MSE of *k*-step-estimators on shrinking neighborhoods. In preparation.