Higher Order Optimal Influence Curves

Peter Ruckdeschel

UNIVERSITÄT BAYREUTH

Mathematisches Institut

Peter.Ruckdeschel@uni-bayreuth.de

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

ICORS 2007

Buenos Aires, September 19, 2007

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:

asymptotically linear estimators:

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:

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

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 (HC-1), 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*

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

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 (HC-1), 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*

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

Limitations of First Order Approach

- So far: asymptotics is of first order, for both ALE and MSE
- Limitations (not a topic today): 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?
  - which construction (achieving an optimally–robust IC asymptotically) to take
- Questions for this talk:

(Q1) *Can we enhance finite sample performance using refined asymptotics?*

(Q2) *Hampel's conjecture:*
—with regard to the corners of (first-order) MSE solution—
Should not a finitely optimal IC be smooth?

Limitations of First Order Approach

- So far: asymptotics is of first order, for both ALE and MSE
- Limitations (not a topic today): 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?
  - which construction (achieving an optimally–robust IC asymptotically) to take
- Questions for this talk:

(Q1) *Can we enhance finite sample performance using refined asymptotics?*

(Q2) *Hampel's conjecture:*
—with regard to the corners of (first-order) MSE solution—
Should not a finitely optimal IC be smooth?

Does first order optimality imply second order optimality?

Classical Optimality (of IC of MLE):

- first order setup:
  - risk-independence in Asympt. Convolution Theorem / for all "bowl-shaped" risks in Asympt. Minimax Theorem
- second order setup: *Pfanzagl*'s catchword
  "First order optimality implies second order optimality"

Robust Optimality (of ICs from class HC-1):

- first order setup (R.& Rieder [& Kohl] (2004/2007))
  - risk-independence of the class
  - risk-dependence of the member within HC-1
  - radius-minimax ICs: risk-independence of the optimal member for all "homogeneous" risks
- second order setup:

(Q3) Does Pfanzagl's catchword apply to the robust setup, and if so in which way?

Does first order optimality imply second order optimality?

Classical Optimality (of IC of MLE):

- first order setup:
  - risk-independence in Asympt. Convolution Theorem / for all "bowl-shaped" risks in Asympt. Minimax Theorem
- second order setup: *Pfanzagl*'s catchword
  "First order optimality implies second order optimality"

Robust Optimality (of ICs from class HC-1):

- first order setup (R.& Rieder [& Kohl] (2004/2007))
  - risk-independence of the class
  - risk-dependence of the member within HC-1
  - radius-minimax ICs: risk-independence of the optimal member for all "homogeneous" risks
- second order setup:

(Q3) Does Pfanzagl's catchword apply to the robust setup, and if so in which way?

Uniform Expansions of the MSE

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* 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

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

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

Uniform Expansions of the MSE

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* 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

$$\mathrm{maxMSE}(S_n) := n \sup_{Q_n \in \tilde{U}_c(r)} \mathrm{MSE}(S_n)$$

$$= \boxed{A_0 + \frac{r}{\sqrt{n}} A_1 + \frac{1}{n} A_2 + \mathrm{o}(\tfrac{1}{n})}$$

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

Second Order Optimality - Symmetric Case

Corollary

Let P_θ *and* ψ *be symmetric:*

Then $A_1 = 2r^2b^2 + v_0^2 + b^2$
i.e., a convex and isotone function in $\|\eta\|_{L_2}$ *and* $\|\eta\|_{L_\infty}$ *— the same terms arising in first order term* A_0.

Consequence:

(ad Q3) Pfanzagl's "rule" for class HC-1:

Second order optimal (s-o-o) IC is of HC-1-form

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

but with adjusted s-o-o clipping height c_1 determined as

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

Second Order Optimality - Symmetric Case

Corollary

Let *P_θ and ψ be symmetric:*

Then $A_1 = 2r^2b^2 + v_0^2 + b^2$
i.e., a convex and isotone function in $\|\eta\|_{L_2}$ and $\|\eta\|_{L_\infty}$ — the same terms arising in first order term A_0.

Consequence:
(ad Q3) Pfanzagl's "rule" for **class** HC-1:
Second order optimal (s-o-o) IC is of HC-1-form

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

but with adjusted s-o-o clipping height c_1 determined as

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

Second Order Optimal Clipping

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

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

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

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

- as c_1 is optimal, s-o risk 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 $\mathrm{O}(1/n^2)$

(ad Q1) there is —albeit little— enhancement by higher order asymptotics

Second Order Optimal Clipping

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, s-o risk 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 $\mathrm{O}(1/n^2)$

(ad Q1) there is —albeit little— enhancement by higher order asymptotics

Optimal c's and corresp. (num.) exact maxMSE at $\mathcal{N}(\theta, 1)$

—$n = 20$, $r = 0.3$—:

		exact risk:		asymptotic risk:		
	c	$\mathrm{relMSE}_n^{\mathrm{ex}}$	$\mathrm{maxMSE}_n^{\mathrm{ex}}$	A_0	$A_0 + \frac{r}{\sqrt{n}} A_1$	$A_0 + \frac{r}{\sqrt{n}} A_1 + \frac{1}{n} A_2$
Median	0+	16.413%	1.911	1.712	1.942	1.875
η_{c_0}	1.213	1.548%	1.667	1.290	1.556	1.615
η_{c_1}	1.017	0.117%	1.643	1.299	1.544	1.596
η_{c_2}	0.972	0.017%	1.642	1.299	1.544	1.596
$\eta_{c_{\mathrm{FZY}}}$	0.991	0.049%	1.642	1.301	1.545	1.596
$\eta_{c_{\mathrm{ex}}}$	0.939	—	1.641	1.307	1.545	1.596

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 in HC-1
c_{FZY}	num. optimization of a proposal by Fraiman et al.
c_{ex}	num. optimization of the (num.) exact MSE

One dimensional Scale

Corollary (Second order optimality for one-dim. scale)

Let *S_n be two-step estimator to IC η_θ (with e.g. MAD as starting estimator)*

Then

$$\mathrm{maxMSE}(S_n) = A_0 + \frac{r}{\sqrt{n}} A_1 + \mathrm{o}(\tfrac{1}{\sqrt{n}})$$

$$\text{for } A_0 = v_0^2 + r^2 b^2, \qquad v_0^2 = \mathrm{E}_\theta\, \eta_\theta^2, \quad b = \sup |\eta_\theta|$$

$$A_1 = v_0^2 + b^2(1+2r^2) + b\Big| l_2\,(3v_0^2 + r^2 b^2) + 2\,v_1 \Big|$$

$$\text{for } l_2 = \frac{d^2}{dt^2}\,\mathrm{E}_\theta\,\eta_t\Big|_{t=\theta}, \quad v_1 = \frac{d}{dt}\,\mathrm{E}_\theta\,\eta_t^2\,\Big|_{t=\theta}$$

Shifting differentiation to P_θ — integration by parts and scale invariance:

$$A_1 = \mathrm{E}\,\eta^2 + b^2(1+2r^2) + \\ + b\Big|\mathrm{E}\,\eta^2(4-2\Lambda) + \mathrm{E}\,\eta L_2(3\,\mathrm{E}\,\eta^2 + r^2 b^2)\Big|$$

$$\text{for} \quad \Lambda = \frac{\partial}{\partial\theta}\log p_\theta, \qquad L_2 = \frac{\partial^2}{\partial\theta^2}\log p_\theta$$

One dimensional Scale

Corollary (Second order optimality for one-dim. scale)

Let *S_n be two-step estimator to IC η_θ (with e.g. MAD as starting estimator)*

Then

$$\mathrm{maxMSE}(S_n) = A_0 + \frac{r}{\sqrt{n}} A_1 + \mathrm{o}(\tfrac{1}{\sqrt{n}})$$

$$\text{for } A_0 = v_0^2 + r^2 b^2, \qquad v_0^2 = \mathrm{E}_\theta\, \eta_\theta^2, \quad b = \sup|\eta_\theta|$$

$$A_1 = v_0^2 + b^2(1+2r^2) + b\Big|l_2\,(3v_0^2 + r^2b^2) + 2\,v_1\Big|$$

$$\text{for } l_2 = \frac{d^2}{dt^2}\mathrm{E}_\theta\,\eta_t\Big|_{t=\theta}, \quad v_1 = \frac{d}{dt}\mathrm{E}_\theta\,\eta_t^2\Big|_{t=\theta}$$

Shifting differentiation to P_θ — integration by parts and scale invariance:

$$A_1 = \mathrm{E}\,\eta^2 + b^2(1+2r^2) + \\ + b\Big|\mathrm{E}\,\eta^2(4-2\Lambda) + \mathrm{E}\,\eta L_2(3\,\mathrm{E}\,\eta^2 + r^2b^2)\Big|$$

$$\text{for } \Lambda = \frac{\partial}{\partial\theta}\log p_\theta, \qquad L_2 = \frac{\partial^2}{\partial\theta^2}\log p_\theta$$

One dimensional Scale

Corollary (Second order optimality for one-dim. scale)

Let *S_n be two-step estimator to IC η_θ (with e.g. MAD as starting estimator)*

Then

$$\begin{aligned} \mathrm{maxMSE}(S_n) &= A_0 + \frac{r}{\sqrt{n}} A_1 + \mathrm{o}(\tfrac{1}{\sqrt{n}}) \\ \text{for } A_0 &= v_0^2 + r^2 b^2, \qquad v_0^2 = \mathrm{E}_\theta\, \eta_\theta^2, \quad b = \sup|\eta_\theta| \\ A_1 &= v_0^2 + b^2(1+2r^2) + b\Big| l_2\,(3v_0^2 + r^2 b^2) + 2\,v_1\Big| \\ &\qquad \text{for } l_2 = \frac{d^2}{dt^2}\mathrm{E}_\theta\, \eta_t\Big|_{t=\theta}, \quad v_1 = \frac{d}{dt}\mathrm{E}_\theta\, \eta_t^2\Big|_{t=\theta} \end{aligned}$$

Shifting differentiation to P_θ — integration by parts and scale invariance:

$$\begin{aligned} A_1 &= \mathrm{E}\,\eta^2 + b^2(1+2r^2) + \\ &\quad + b\Big|\mathrm{E}\,\eta^2(4-2\Lambda) + \mathrm{E}\,\eta L_2(3\,\mathrm{E}\,\eta^2 + r^2 b^2)\Big| \\ \text{for} \quad \Lambda &= \frac{\partial}{\partial\theta}\log p_\theta, \qquad L_2 = \frac{\partial^2}{\partial\theta^2}\log p_\theta \end{aligned}$$

Second Order Optimality Problems

MSE-2 $\quad F_n(\eta) := A_0(\eta) + \frac{r}{\sqrt{n}} A_1(\eta) = \min!$ $\quad \eta$ IC, $\eta \in L_3(P)$

Structure of the problem suited for convex optimization

- admitted functions form convex set
- F_n is coercive $\rightsquigarrow$ restriction to some bounded L_∞-ball possible
- eventually in n, F_n is weakly lower semicontinuous in L_3 and strictly convex $\Longrightarrow$ unique minimum solution exists
- Slater condition fulfilled $\rightsquigarrow$ Lagrange multipliers exist
- equivalence to Hampel-problem: for $r_n = \frac{r}{\sqrt{n+r}}$ and

$$H_n(\eta) = \mathrm{E}\,\eta^2 + r_n b\Big(2\,\mathrm{E}\,\eta^2(2-\Lambda) + \mathrm{E}\,\eta L_2(3\,\mathrm{E}\,\eta^2 + r^2 b^2)\Big)$$

HP-2 $\quad H_n(\eta) = \min!$ $\quad$ for η IC $\quad$ and $\sup|\eta| \le b$

Second Order Optimality Problems

MSE-2 $$F_n(\eta) := A_0(\eta) + \frac{r}{\sqrt{n}} A_1(\eta) = \min ! \qquad \eta \text{ IC}, \; \eta \in L_3(P)$$

Structure of the problem suited for convex optimization

- admitted functions form convex set
- F_n is coercive $\rightsquigarrow$ restriction to some bounded L_∞-ball possible
- eventually in n, F_n is weakly lower semicontinuous in L_3 and strictly convex $\Longrightarrow$ unique minimum solution exists
- Slater condition fulfilled $\rightsquigarrow$ Lagrange multipliers exist
- equivalence to Hampel-problem: for $r_n = \frac{r}{\sqrt{n+r}}$ and

$$H_n(\eta) = \mathrm{E}\,\eta^2 + r_n b\Big(2\,\mathrm{E}\,\eta^2(2-\Lambda) + \mathrm{E}\,\eta L_2(3\,\mathrm{E}\,\eta^2 + r^2 b^2)\Big)$$

HP-2 $$H_n(\eta) = \min ! \qquad \text{for } \eta \text{ IC} \quad \text{and} \;\; \sup|\eta| \le b$$

Solution for fixed b

Solution to HP-1 for fixed b
"Hampel-type-1" (HC-1)

$$\hat{\eta} = Y \min\{1, \frac{b}{|Y|}\}$$

for

$$Y = A\Lambda - a$$

with scores Λ, Lagrange multipliers A, a, and bias bound b

Gaussian case:

$$Y = Ax^2 - a$$

Solution for fixed b

Solution to HP-2 for fixed b

"Hampel-type-2" (HC-2)

$$\hat\eta = Y_n \min\{1, \frac{b}{|Y_n|}\}, \qquad Y_n = \frac{Y - r_n b L_2 (r^2 b^2 + 3 v_0^2)/2}{1 + r_n b(4 - 2\Lambda + 3 l_2)}$$

for

$$Y = A\Lambda - a, \qquad v_0^2 = \mathrm{E}\, Y^2, \qquad l_2 = \mathrm{E}\, L_2$$

with scores Λ, Lagrange multipliers A, a, and bias bound b, and second order radius term $r_n = r/(\sqrt{n} + r)$

Gaussian case:

$$Y_n = \frac{Ax^2 - a - r_n b(x^4 - 5x^2 + 2)(3v_0^2 + r^2 b^2)/2}{1 + r_n b(6 - 2x^2 + 3 l_2)}$$

Solution for fixed b

Solution to HP-2 for fixed b

"Hampel-type-2" (HC-2)

$$\hat{\eta} = Y_n \min\{1, \frac{b}{|Y_n|}\}, \qquad Y_n = \frac{Y - r_n b L_2 (r^2 b^2 + 3v_0^2)/2}{1 + r_n b(4 - 2\Lambda + 3 l_2)}$$

for

$$Y = A\Lambda - a, \qquad v_0^2 = \mathrm{E}\, Y^2, \qquad l_2 = \mathrm{E}\, L_2$$

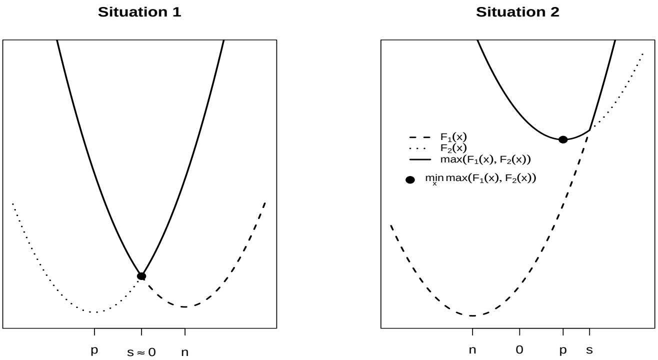
Gaussian case:

$$Y_n = \frac{Ax^2 - a - r_n b(x^4 - 5x^2 + 2)(3v_0^2 + r^2 b^2)/2}{1 + r_n b(6 - 2x^2 + 3 l_2)}$$

Problem:

- $\left| \quad \cdot \quad \right|$-expression in s-o-term A_1 in maxMSE:
  b in f-o-term A_0 may be induced by positive or negative bias
- i.e.; $\mathrm{maxMSE}(\eta) = A_0(\eta) + r/\sqrt{n} \max\big(A_1(\eta, -b), A_1(\eta, b)\big)$

Positive or Negative Bias?



- to be checked for any second-order MSE-solution
- numerically for Gaussian scale case:
  left situation (optimum in intersection point of parabolas)

Hampel-type ICs Second Order Optimal?

(ad Q3) Possible gain in (s-o-)maxMSE w.r.t. s-o-clipping-adjusted HC-1-type IC $< 10^{-5}$!!
—consequence:

- may stay in class HC-1 of Hampel-type ICs
- simply adjust clipping height w.r.t. first order optimal solution
- Pfanzagl's "rule" for **class** HC-1

(ad Q2) General feature:

- no matter whether optimal solution $\hat{\eta}$ is of type HC-1 or HC-2:

Solution involves clipping!

- if $Y'_{[n]} \neq 0$ in clipping points: *non-smooth* optimal IC
- argument applies to arbitrary asymptotic order (3rd, 4th,...)

Hampel-type ICs Second Order Optimal?

(ad Q3) Possible gain in (s-o-)maxMSE w.r.t. s-o-clipping-adjusted HC-1-type IC $< 10^{-5}$!!
—consequence:

- may stay in class HC-1 of Hampel-type ICs
- simply adjust clipping height w.r.t. first order optimal solution
- Pfanzagl's "rule" for **class** HC-1

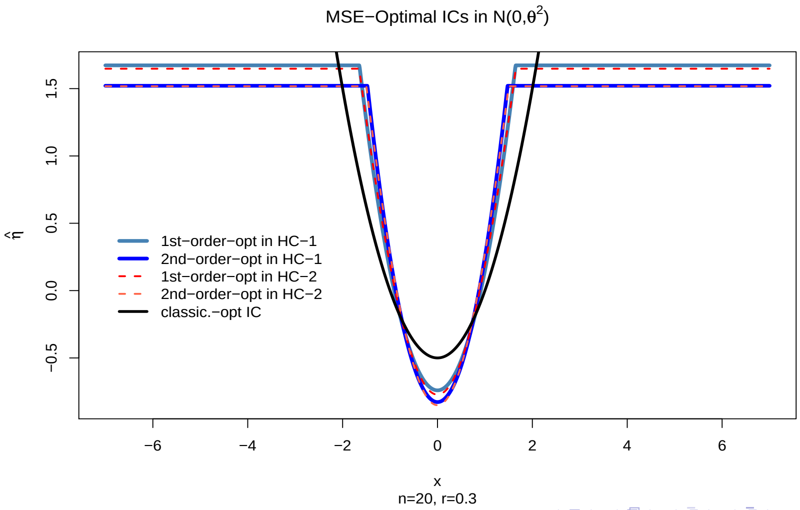
(ad Q2) General feature:

- no matter whether optimal solution $\hat{\eta}$ is of type HC-1 or HC-2:

Solution involves clipping!

- if $Y'_{[n]} \neq 0$ in clipping points: *non-smooth* optimal IC
- argument applies to arbitrary asymptotic order (3rd, 4th,...)

Second Order MSE-Optimal IC to $r = 0.3$



Optimal b's and corresp. empirical maxMSE

emp. results for $\mathcal{N}(0, \theta^2)$ at $n = 20$, $r = 0.3$ at $M = 90000$ runs:

	b	empirical risk: $\mathrm{relMSE}_n^{\mathrm{sim}}$	$\mathrm{maxMSE}_n^{\mathrm{sim}}$		asymptotic risk: A_0	$A_0 + \frac{r}{\sqrt{n}} A_1$
MAD	1.166	24.18%	1.822	[1.801; 1.843]	1.223	1.487
η_{b_0}	1.671	27.48%	1.870	[1.836; 1.905]	0.892	1.075
η_{b_1} (HC-1)	1.530	8.75%	1.596	[1.572; 1.619]	0.905	1.057
η_{b_1} (HC-2)	1.531	8.63%	1.594	[1.571; 1.617]	0.906	1.060
$\eta_{b_{\mathrm{sim}}}$	1.346	—	1.467	[1.450; 1.484]	0.945	1.105

b_0	f-o-o: optimized A_0 within HC-1
b_1	s-o-o: optimized $A_0 + \frac{r}{\sqrt{n}} A_1$ within HC-1/HC-2
b_{sim}	num. optimization of the (empirical) maxMSE within HC-1

Consequences:

- first order asymptotics too optimistic

(ad Q1) considerable enhancement by 2nd order asymptotics —

- but: still room for improvement by 3rd order asymptotics

One dimensional Location and Scale for symmetric F

Corollary (Second order optimality for one-dim. location and scale)

Let *S_n be two-step estimator to IC η_θ (with e.g. (Median,MAD) as starting estimator)*

Then

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

for $A_0 = \mathrm{E}_\theta\, |\eta_\theta|^2 + r^2 b_\theta^2,\; b_\theta = \sup |\eta_\theta|$ *and*

A_1 *only slightly more complicated than in pure scale case*

Structure of the Solution

- similar arguments as in scale case
- location component is odd, scale component even
- adaptivity also holds for second order asymptotics (but nuisance part has to have bounded IC)
- positive/negative bias: here right situation in the parabola picture (optimum in a vertex of a parabola)

(ad Q3) possible gain in (s-o-)maxMSE w.r.t. s-o-clipping-adjusted HC-1 IC $\ll 1\%$!! —hence grossly speaking:

- may stay in class HC-1 of Hampel-type ICs
- simply adjust clipping height w.r.t. first order optimal solution
- again: Pfanzagl's "rule" for **class** HC-1

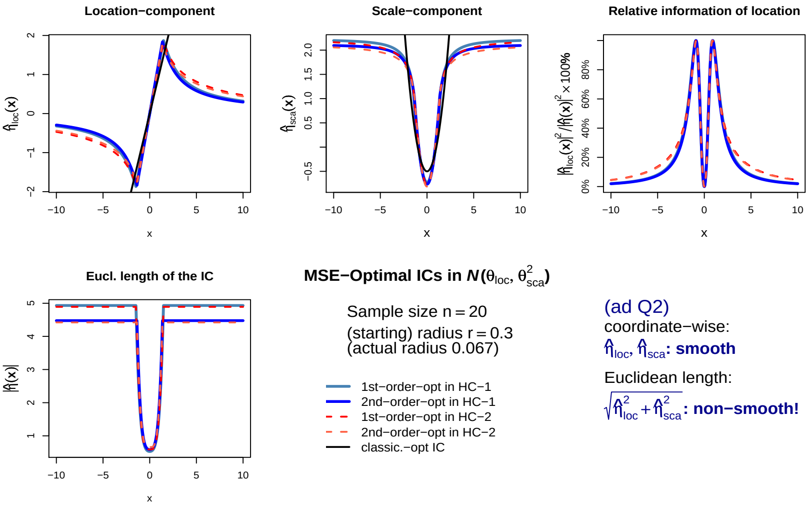
Structure of the Solution

- similar arguments as in scale case
- location component is odd, scale component even
- adaptivity also holds for second order asymptotics (but nuisance part has to have bounded IC)
- positive/negative bias: here right situation in the parabola picture (optimum in a vertex of a parabola)

(ad Q3) possible gain in (s-o-)maxMSE w.r.t. s-o-clipping-adjusted HC-1 IC $\ll 1\%$!! —hence grossly speaking:

- may stay in class HC-1 of Hampel-type ICs
- simply adjust clipping height w.r.t. first order optimal solution
- again: Pfanzagl's "rule" for **class** HC-1

Second Order MSE-Optimal IC to $r = 0.3$



Optimal b's and corresp. empirical maxMSE

emp. results for $\mathcal{N}(\theta_{\rm loc}, \theta^2_{\rm sca})$ at $n = 20$, $r = 0.3$ at $M = 90000$ runs:

		empirical risk:				asymptotic risk:	
	b	$\mathrm{relMSE}^{\rm sim}_n$	$\mathrm{maxMSE}^{\rm sim}_n$			A_0	$A_0 + \frac{r}{\sqrt{n}} A_1$
(Median;MAD)	1.713	22.38%	3.747	[3.716;	3.777]	3.057	3.629
η_{b_0}	2.221	37.37%	4.205	[4.117;	4.294]	2.154	2.768
η_{b_1} (HC-1)	2.116	23.52%	3.782	[3.713;	3.850]	2.161	2.757
η_{b_1} (HC-2)	2.103	19.51%	3.659	[3.590;	3.720]	2.167	2.753
$\eta_{b_{\rm sim}}$	1.744	—	3.061	[3.033;	3.090]	2.406	3.069

b_0	f-o-o: optimized A_0 within HC-1
b_1	s-o-o: optimized $A_0 + \frac{r}{\sqrt{n}} A_1$ within HC-1/HC-2
$b_{\rm sim}$	num. optimization of the (empirical) maxMSE within HC-1

Consequences:

- again: first order asymptotics too optimistic

(ad Q1) again: enhancement by 2nd order asymptotics —

- but even 2nd order asymptotics probably not enough

Summary: Answers to (Q1)-(Q3)

(Q1) *Can we enhance finite sample performance using refined asymptotics?*
Yes, we can — for location only a little, for scale and location/scale considerably...

(Q2) Hampel's conjecture: *"Should not a finitely optimal IC be smooth?"*
Regarding higher order asymptotics: **No**, they should not.

(Q3) Does Pfanzagl's catchword *"First order optimality implies second order optimality"* apply to the robust setup, and if so in which way? Grossly speaking:
Yes, it does **classwise** for class (HC-1). However, first order optimal clipping height is **too optimistic**.

Bibliography

- H. Rieder (1994): *Robust Asymptotic Statistics. Springer.*
- P. Ruckdeschel (2005(a)): Higher order asymptotics for the MSE of M estimators on shrinking neighborhoods. Submitted. 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 the median on shrinking neighborhoods. Submitted. Also available in http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/medmse.pdf.
- P. Ruckdeschel (2005(c)): Higher order asymptotics for the MSE of k-step-estimators on shrinking neighborhoods. In preparation.

For references please confer the handout to this talk on my web-page.

Thank you for your attention!

Uniform Expansions of the MSE II

Exact expressions for term A_1 for 1-step-estimator

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

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

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

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

Uniform Expansions of the MSE II

Exact expressions for term A_1 for 1-step-estimator

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

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

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

M-est put $\tilde{\eta}_\theta = \eta_\theta$

Uniform Expansions of the MSE II

Exact expressions for term A_1 for 1-step-estimator

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

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

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 $\tilde\eta_\theta = \eta_\theta$

Outlook: Total Variation Neighborhoods

PhD project of M. Brandl; preliminary results:

- total variation $\hat{=}$ replacement outliers
- maxMSE has a higher order expansion
- asymmetric case
  - already in first order asymptotics different solutions for convex contamination and total variation [Ri:94]
  - asymmetric clipping for first order optimal optimal solution [Ri:94]
- symmetric case
  - $A_1 = 0$ — first correction term in $\max\mathrm{MSE}$ of order $\mathrm{O}(n^{-1})$

$\Longrightarrow$ faster convergence of $\max\mathrm{MSE}$

(ad Q3) Pfanzagl's "rule" memberwise in HC-1

(ad Q1) no enhancement by second order asymptotics

Outlook: Total Variation Neighborhoods

PhD project of M. Brandl; preliminary results:

- total variation $\hat{=}$ replacement outliers
- maxMSE has a higher order expansion
- asymmetric case
  - already in first order asymptotics different solutions for convex contamination and total variation [Ri:94]
  - asymmetric clipping for first order optimal optimal solution [Ri:94]
- symmetric case
  - $A_1 = 0$ — first correction term in maxMSE of order $\mathrm{O}(n^{-1})$

$\Longrightarrow$ faster convergence of maxMSE

(ad Q3) Pfanzagl's "rule" memberwise in HC-1

(ad Q1) no enhancement by second order asymptotics

Outlook: Total Variation Neighborhoods

PhD project of M. Brandl; preliminary results:

- total variation $\hat{=}$ replacement outliers
- maxMSE has a higher order expansion
- asymmetric case
  - already in first order asymptotics different solutions for convex contamination and total variation [Ri:94]
  - asymmetric clipping for first order optimal optimal solution [Ri:94]
- symmetric case
  - $A_1 = 0$ — first correction term in maxMSE of order $\mathrm{O}(n^{-1})$

$\Longrightarrow$ faster convergence of maxMSE

(ad Q3) Pfanzagl's "rule" memberwise in HC-1

(ad Q1) no enhancement by second order asymptotics