

Höhere Ordnungsasymptotik gleichmäßig auf Umgebungen und deren Konsequenzen für optimal-robuste Schätzverfahren (wissenschaftlicher Vortrag im Rahmen der Habilitation)

Peter Ruckdeschel

peter.ruckdeschel@itwm.fraunhofer.de
http://www2.itwm.fhg.de/de/fm_employees__Peter_Ruckdeschel/ruckdeschel/

4. Mai 2011

Motivation: Robustheit gegen Ausreißer

Motivation: ein elementares Beispiel

Ausreißer und Umgebungen

Konzepte der robusten Statistik: Funktionale

Stabilitätskonzepte der robusten Statistik

Inhalt

Motivation: Robustheit gegen Ausreißer

Asymptotik erster Ordnung in robuster Statistik – ein Abriss

Höhere Ordnungsasymptotik gleichmäßig auf Umgebungen

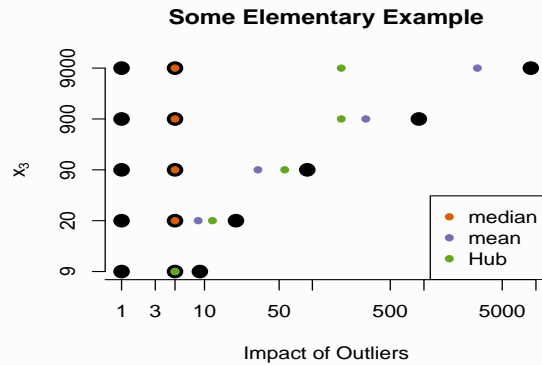
Spezialfall einer Dimension: Höhere-Ordnungs-Optimalität

Vergleich k -Schritt- und M -Schätzer

Motivation: ein elementares Beispiel

- ▶ Gaußsches Lokationsproblem: $X_i = \theta + \varepsilon_i$, $\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$
- ▶ Ziel: schätze θ mittels X_i
- ▶ Kandidaten
$$S_1(x_1, \dots, x_n) = \bar{X}_n \text{ (MLE)}$$
$$S_2(x_1, \dots, x_n) = \text{median}(X_i)$$
$$S_3(x_1, \dots, x_n) = \text{Hub}_c(X_i)$$
- ▶ Zahlenbeispiel: $n = 3$, $x_1 = 1$, $x_2 = 5$, $x_3 = 9$
- ▶ was passiert, wenn $x_3 = 20, 90, 900, 9000$?

Motivation: ein elementares Beispiel (Forts.)



Problem: wenige Ausreißer machen Mittelwert unbrauchbar

Ausreißer und Umgebungen

- ▶ Was macht eine Beobachtung zu einem Ausreißer?
 - ▶ betrifft nur **wenige** Beobachtungen (5%–10%)
 - ▶ **unkontrollierbar**, aus **unbekannter** Vtlg, **unvorhersehbar**
 - ▶ meistens: nicht mit Sicherheit von id. Beob. trennbar
 - ▶ Ausreißerverteilung **variiert** von Beob. zu Beob.
- ▶ Trotzdem: reale Stichprobe sollte **nah** an der idealen Situation liegen
- ▶ **nah in welchem Sinn?**
 - ▶ im Sinn von **Momenten**? schlechte Idee!
 - ▶ im Sinn der **schwachen Konvergenz** / von Zentralem Grenzwertsatz
- ▶ **Gross Error:** $\mathcal{U}_c(F, \varepsilon) = \{G \mid \exists H: G = (1 - \varepsilon)F + \varepsilon H\}$
 interpretierbar als $\{\mathcal{L}(X^{\text{re}}) \mid X^{\text{re}} = (1 - U)X^{\text{id}} + UX^{\text{di}}\}$,
 wobei $U \sim \text{Bin}(1, \varepsilon)$, $X^{\text{id}} \sim F$, $X^{\text{di}} \sim H$, und $U, X^{\text{id}}, X^{\text{di}}$ unabh.

Konzepte der robusten Statistik: Funktionale

- ▶ parametrische Verfahren meist interpretierbar als **Funktionale** T , die **Verteilungen** in **Parameter** abbilden
- ▶ kanonische Forderung: Fisher-Konsistenz, i.e.; $T(F_\theta) = \theta$
- ▶ Schätzer entsteht durch Anwendung von T auf emp. Verteilungsfunktion $S_n = T(\hat{F}_n)$
- ▶ Beispiele
 - ▶ **mean:** $\bar{X}_n = T_1(\hat{F}_n) = \int x \hat{F}_n(dx)$, i.e. $T_1(F) = \int x F(dx)$
 - ▶ **median:** $T_2(\hat{F}_n) = \hat{F}_n^-(1/2)$, i.e. $T_2(F) = F^-(1/2)$
 - ▶ **MLE:** θ_n^{MLE} Maximum von $\vartheta \mapsto \sum \log p_\vartheta(x_i)$ bzw. von $\vartheta \mapsto \int \log p_\vartheta(x) \hat{F}_n(dx)$, i.e.
 $T_{\text{MLE}}(F) = \text{argmax}_\vartheta \left(\int \log p_\vartheta(x) F(dx) \right)$

Stabilitätskonzepte der robusten Statistik

Qualitative Robustheit aka Stetigkeit

- ▶ T ist *qualitativ robust* in F wenn T glm. schwach stetig in n ist, i.e. $\forall \varepsilon \exists \delta, N$ s.d. $\forall G \in \mathcal{U}_\pi(F, \delta)$ und $\forall n > N$
 $d_\pi(T \circ \hat{F}_n, T \circ \hat{G}_n) \leq \varepsilon$.

Bruchpunkt aka nächstgelegene Singularität

- ▶ T hat Bruchpunkt ε at F wenn für jedes $\varepsilon' < \varepsilon$ gilt
 $\sup_{G \in \mathcal{U}(F, \varepsilon')} |T(G) - \theta| < \infty$

Influenzkurve (IC) aka Ableitung

- ▶ T hat IC ψ wenn ψ in Richtung eines Diracmaßes Gâteaux-diff'bar ist i.e.
 $\psi(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} (T((1 - \varepsilon)F + \varepsilon I_{\{x\}}) - T(F))$

Bücher: Huber(1981), Hampel et al.(1986), Maronna et al.(2006)

Robustheitsbegriffe im Beispiel

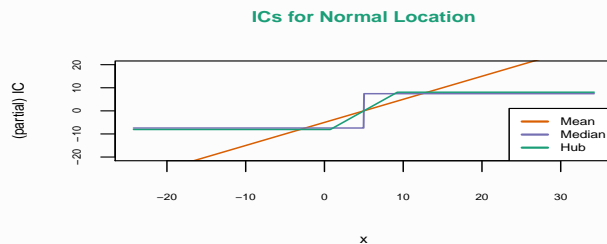
Qualitative Robustheit aka Stetigkeit

- ▶ Mittelwert ist **nicht** (qualitativ) robust.
- ▶ Median und Hub sind qualitativ robust.

Bruchpunkt aka nächstgelegene Singularität

- ▶ Mittelwert hat (minimalen) Bruchpunkt $1/n$.
- ▶ Median und Hub haben (maximalen) Bruchpunkt $1/2$.

Influenzkurve (IC) aka Ableitung



Asymptotik erster Ordnung in robuster Statistik – ein Abriss

Idealer Setup: L_2 -differenzierbares Modell
 Influenzkurven und asymptotisch lineare Schätzer
 Schrumpfende Umgebungen $\mathcal{U}(P_\theta, r)$ zu Radius r
 Optimal-robuste Schätzer

L_2 -differenzierbares Modell

$\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^k$ offen

- ▶ Beispiel: Gaußsche Lokation: $\mathcal{P} = \{\mathcal{N}(\theta, 1) \mid \theta \in \Theta\}$, $\Theta = \mathbb{R}$

Glattheit: L_2 -Differenzierbarkeit

- ▶ $\sqrt{p_{\theta+h}} = \sqrt{p_\theta} \left(1 + \frac{1}{2} \Lambda_\theta^\top h\right) + o(|h|)$ in L_2
- ▶ Fisher-Information $\mathcal{I}_\theta := \int \Lambda_\theta \Lambda_\theta^\top dP_\theta$ endlich und regulär

Konsequenz:

- ▶ $P_{\theta+h/\sqrt{n}}^n$ und P_θ^n sind benachbart
- ▶ $\log dP_{\theta+h/\sqrt{n}}^n / dP_\theta^n = \frac{1}{\sqrt{n}} \sum_i h^\top \Lambda_\theta(x_i) - \frac{1}{2} h^\top \mathcal{I}_\theta h + o_{P_\theta^n}(1)$

⇒ Modell ist LAN (lokal asymptotisch normal)

Influenzkurven (ICs) und ALEs

Influenzkurve (IC)

$$\psi_\theta \in L_2^k(P_\theta) \quad \text{s.d.} \quad \mathbb{E}_\theta \psi_\theta = 0, \quad \mathbb{E}_\theta \psi_\theta \Lambda_\theta^\top = \mathbb{I} \quad (\mathbb{E}_\theta = \mathbb{E}_{P_\theta})$$

hier: IC als mögliche Linearisierung eines Schätzers

Asymptotisch lineare Schätzer (ALE): besitzen eine Entwicklung

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

für eine gewisse IC ψ_θ

- ▶ (1) \iff lokal glm. as. Normalität/Fisher-Konsistenz

Beispiele

- ▶ in \mathcal{P} : $S_n = \bar{X}_n$ $- \psi_\theta(x) = x - \theta$,
- ▶ $S_n = \text{Median}_n$ $- \psi_\theta(x) = \sqrt{\frac{\pi}{2}} \text{sign}(x - \theta)$

Schrumpfende Umgebungen $\mathcal{U}(P_\theta, r)$ zu Radius r

- ▶ für ALEs wichtige as. Eigenschaften ablesbar an IC ψ :
 - ▶ *Gross error sensitivity* GES ist $\sup_x |\psi|$
 - ▶ *as. Var* ist $E \psi \psi^\tau$
- ⇒ (as. Cramér-Rao): optimale IC im idealen Modell ist $\mathcal{I}^{-1}\Lambda$
- ▶ Skalierung für fixen Umgebungsradius ε :
 - Bias mit $O(\varepsilon)$ und Varianz mit $O(1/n)$,d.h. 16 Ausreißer bei $n = 100 \hat{=} 8$ (nicht 4) Ausreißer bei $n = 25$
- ▶ also für wachsendes n Bias dominant; dann aber auch (as. fehlerfrei) detektierbar (vgl. P.R.[06])
- ⇒ Schrumpfen mit $\varepsilon = \varepsilon_n = r/\sqrt{n}$

Höhere Ordnungsasymptotik gleichmäßig auf Umgebungen

Grenzen des Erstordnungs-Ansatzes und offene Fragen
verschiedene Konstruktionen zur selben IC
Ein Exkurs zur Genese. . . : Lausanne 2003
Bisherige Ansätze
Gleichgradige Integrierbarkeit
Gleichmäßige Entwicklungen des MSE

MSE-optimale IC

- ▶ für schrumpfende Umgebungen \mathcal{U} gilt:
 - ▶ *max.as.Bias* auf \mathcal{U} ist $r \sup |\psi|$ und
 - ▶ *max.as.MSE* auf \mathcal{U} ist $\text{asMSE}(\psi, r) = E |\psi|^2 + r^2 \sup |\psi|^2$
- ↔ robuste Optimalität zurückgeführt auf IC;
zweiter (Konstruktions-)Schritt:
finde zu optimaler IC geeigneten ALE

MSE-Problem: finde IC $\hat{\psi}$, die *asMSE* auf \mathcal{U} minimiert

Thm. 5.5.1, Rieder[94]

Die eindeutige Lösung ist eine IC $\tilde{\psi}$ von Hampel-Form, i.e.:

$$\tilde{\psi} = (A\Lambda - a)w \quad w = \min \{1, b/|A\Lambda - a|\}$$

wobei A, a, b gewährleisten dass $E \tilde{\psi} = 0$, $E \tilde{\psi} \Lambda^\tau = \mathbb{I}_k$, und
(MSE) $r^2 b = E (|A\Lambda - a| - b)_+$

offene Fragen im Erstordnungs-Ansatzes

- ▶ Bis jetzt: Asymptotik ist erster Ordnung (ALE + asMSE)
- ▶ offene Fragen
 - (Q1) Einflussgrößen und Geschwindigkeit der Konvergenz?
— wie beeinflussen Radius r , Stichprobenumfang n und Stutzhöhe b die Approximationsgüte quantitativ?
 - (Q2) Welche Konstruktion soll man verwenden?
- ↔ **Höhere Ordnungsasymptotik glm. auf Umgebungen**
 - (Q3) (1.Ord)-opt. ALEs für finites n mit feinerer Asy. verbesserbar?
 - (Q4) (Wie) Überträgt sich Pfanzagls Schlagwort [klass. Kontext]:
"First order optimality implies second order optimality"?
 - (Q5) Theoretische Evidenz für Hampels Vermutung:
"Finitely-optimal IC should be smooth"

zu (Q2): verschiedene Konstruktionen zur selben IC

- ▶ Erstordnungs-Asymptotik kann nicht unterscheiden zwischen

- ▶ M-Schätzer (hängt nicht ab von $\theta_n^{(0)}$!):

$$\theta_n^{(z)} \text{ s.d. } g_n(\theta_n^{(z)}) = 0 \quad \text{für } g_n(\theta) = \sum_{i=1}^n \psi_{\theta}(X_i),$$

- ▶ k -Schritt-Schätzer: für Startschätzer $\theta_n^{(0)}$,

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

Ein Exkurs zur Genese I. . . : Lausanne 2003

Untersuchung von Fragen (Q1) und (Q2) für Median und M-Schätzer in \mathcal{P} auf Basis von Simulationen

- ▶ Übereinstimmung finit \leftrightarrow asy. MSE's
 - ▶ ideales Modell: log-log-Regression ergibt emp. Rate $1/n$
 - ▶ kontam. Modell: emp. Rate $1/\sqrt{n}$
 - ▶ relative MSE's: auch im kontam. Modell emp. Rate $1/n$
- ▶ Konvergenz wird langsamer für wachsende r, b
- ▶ keine großen Unterschiede bei Konstruktionen

Ein Exkurs zur Genese II. . . : Lausanne 2003

- ▶ vorläufige (empirisch ermittelte) Entwicklung: Endergebnis

Satz (für beliebiges n) — R.[10(a)]

- ▶ für den Stichprobenmedian im idealen Modell $\mathcal{N}(0, 1)$

$$n \text{MSE}(\text{Med}_n, \mathcal{N}(0, 1)) = \frac{\pi}{2} \left[1 + \left(\frac{\pi}{2} - 2 \underbrace{\left[\frac{-1}{n} \right]}_{\substack{\text{für mittleren Med.} \\ \text{falls } n \text{ gerade}}} \right) / n \right] + o\left(\frac{1}{n}\right)$$

- ▶ glm. Konv. für Med_n auf Umgeb. um ideales Modell

$$n \sup_{F^{re}} \text{MSE}(\text{Med}_n, F^{re}) = \frac{\pi}{2} (1 + r^2) \left(1 + \frac{2r}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \right)$$

Bisherige Ansätze für Fragen (Q1)-(Q5)

- ▶ *von Mises-(Schätzer-)Entwicklung* (Jurečkova und Koautoren, [83–97])
 - Idee: Für zwei Schätzer S_n, S'_n , entwickle $\Delta_n = S_n - S'_n$ weiter (für glatte ICs)
 - aber Entwicklung muss nicht existieren (z.B. Median);
dann: *Bahadur-Kiefer Darstellung* für das Restglied
– wg. Korrelation: $\mathcal{L}(\Delta_n)$ genügt nicht f. Vergl. $\mathcal{L}(S_n), \mathcal{L}(S'_n)$
- ▶ *Verteilungs-Entwicklung* (Edgeworth- / Sattelpunktnäherung.) (e.g. Ronchetti und Welsh [02])
 - ▶ flexibler aber (S.-approx.) keine expliziten analytischen Terme
 - + bei glgrd. Integrierbarkeit hinreichend für (MSE-)risk

in diesen Ansätzen: keine glm. Aussagen auf Umgebungen

Gleichgradige Integrierbarkeit

- ▶ Verteilungskonvergenz $\not\Rightarrow$ Konvergenz der Momente
- ▶ Vitali&Skorohod: glgrd. Int'barkeit \iff Konv. d. Momente
- ▶ **nicht "gratis"**:
 - ▶ Momente des Median konv. nicht wenn F zu "fette" Flanken
 - ▶ Selbst bei "dünnen" Flanken und $0 < \varepsilon_n \rightarrow 0$ beliebig schnell: Momente des Median unbeschränkt auf Umgebung
- ▶ aber: glgrd. Int'barkeit (in n und $\mathcal{U}(\cdot, \frac{r}{\sqrt{n}})$) erreichbar durch **Modifikation**
 - ▶ der **Umgebungen** (M-Schätzer, Quantile): nur Stichproben mit weniger als (FS-)Bruchpunkt vielen Kontam.'s zugelassen
 - ▶ der **Schätzer**: Null-Setzung außerhalb expon. in n wachsendem Kompaktum; immer möglich: z.B. Cramér-von-Mises Min-Distanz-Schätzer
- ▶ Modifikation **exponentiell vernachlässigbar**
- ▶ Beweismittel: exponentielle Konzentrationsschranken/Höfding

Gleichmäßige Entwicklungen des MSE I

Satz (R. [10(a,b,c),11])

Sei $\theta \mapsto \psi_\theta$ glatt in $L_1(P_\theta)$,

S_n ein M - oder ein k -Schritt-Schätzer zur IC ψ_θ , und

sei der Startschätzer $\theta_n^{(0)}$ zum k -Schritt-Schätzer

- ▶ glm. $n^{1/4+\delta}$ -konsistent auf \mathcal{U} für ein $\delta > 0$
- ▶ glm. quadrat-integrierbar in n und auf \mathcal{U}

Dann

$$\begin{aligned} \max_{Q_n \in \mathcal{U}(r)} \text{MSE}(S_n) &:= n \sup_{Q_n \in \mathcal{U}(r)} \text{MSE}(S_n) \\ &= \boxed{A_0 + \frac{r}{\sqrt{n}} A_1 + \frac{1}{n} A_2 + o\left(\frac{1}{n}\right)} \end{aligned}$$

für $A_0 = E_\theta |\psi_\theta|^2 + r^2 \sup |\psi_\theta|^2$ und A_1, A_2 Konstanten, die abhängen von ψ_θ, r , & f. k -Schritt-Schätzer auch von $\theta_n^{(0)}$

ad (Q1): Konv. bei Kontam. um eine $n^{-1/2}$ -Potenz langsamer

Gleichmäßige Entwicklungen des MSE II

Explizite Terme für A_1 für 1-Schritt-Schätzer in einer Dimension

Sei ψ_θ beschränkt und zwei mal diff'bar in $L_1(P_\theta)$,

$$\theta_n^{(0)} = \theta + \frac{1}{n} \sum \tilde{\psi}_\theta(x_i) + o_{L_1(\mathcal{U})}(n^{-1/2}) \text{ für eine beschr. IC } \tilde{\psi}_\theta,$$

Dann

$$\begin{aligned} A_1 &= 2 \text{Cov}_\theta(\psi_\theta, \tilde{\psi}_\theta) - \text{Var}_\theta \psi_\theta^2 + b_\theta^2 \\ &\quad + 2b_\theta^2 \frac{d}{dt} \text{Cov}_\theta(\psi_t, \tilde{\psi}_\theta) \Big|_{t=\theta} + 2\tilde{b}_\theta^2 \frac{d}{dt} \text{Var}_\theta \psi_t \Big|_{t=\theta} \\ &\quad + \frac{d^2}{dt^2} E_\theta \psi_t \Big|_{t=\theta} \left[b_\theta \text{Var}_\theta \tilde{\psi}_\theta + 2\tilde{b}_\theta \text{Cov}_\theta(\psi_\theta, \tilde{\psi}_\theta) \right] \\ &\quad + r^2 \tilde{b}_\theta b_\theta \left[2 + \tilde{b}_\theta \frac{d^2}{dt^2} E_\theta \psi_t \Big|_{t=\theta} \right] \end{aligned}$$

wobei $b_\theta = \sup |\psi_\theta|$, $\tilde{b}_\theta = \limsup_{\varepsilon \downarrow 0} \sup |\tilde{\psi}_\theta| \mathbb{I}(|\psi_\theta| \geq b_\theta - \varepsilon)$

M-Sch. Setze $\tilde{\psi}_\theta = \psi_\theta$ (Zusatzbedingungen)

Spezialfall einer Dimension: Höhere-Ordnungs-Optimalität

Gleichmäßigkeit unabhängig vom Startschätzer

Kommentare und Beweisskizze

Konsequenz: Zweitordnungsoptimalität

Vergleich der "Optimalitäten"

Gleichmäßige MSE-Entwicklung für M-Schätzer und Median im Lokationsmodell

Satz (R. [10(a,b)])

Sei ψ_θ beschränkte, monotone IC s.d. $\theta \mapsto E_\theta \psi_\theta^j$ glatt

P_θ habe höchstens polynomiale Flanken

Sei S_n ein M-Schätzer zu ψ_θ

$$\begin{aligned} \text{Dann} \quad \max \text{MSE}(S_n) &:= n \sup_{Q_n \in \mathcal{U}(r)} \text{MSE}(S_n) \\ &= A_0 + \frac{r}{\sqrt{n}} A_1 + \frac{1}{n} A_2 + o\left(\frac{1}{n}\right) \end{aligned}$$

für $A_0 = v_0^2 + r^2 b^2 = E_\theta |\psi_\theta|^2 + r^2 \sup |\psi_\theta|^2$ und Konstanten A_1, A_2 die abhängen von ψ_θ und r

$\max \text{MSE}$ wird bereits erreicht, wenn Kontamination rechts [links] v. $\psi_\theta^{-1}(\sup |\psi_\theta|) \pm b\sqrt{2 \log(n)/n}$ konzentriert ist.

Konsequenz: Zweitordnungsoptimalität

Korollar Sei P_θ und ψ symmetrisch:

Dann $A_1 = 2r^2 b^2 + v_0^2 + b^2$
und (2.Ord.)Max-Risk ist $R_1(S_n) = r^2 b^2 + v_0^2 + \frac{r}{\sqrt{n}} A_1$

Also 2.Ord.-optimale IC ist von Hampelgestalt $A \wedge \min\{1, c_1/|\Lambda|\}$ mit 2.Ord.-opt. Stützhöhe c_1 gegeben durch

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

- Es gilt stets: $c_1 < c_0$, d.h.
1.Ord.-Asymptotik ist generell zu optimistisch
- Weil c_1 optimal, verhält sich R_1 lokal wie Parabel mit Scheitel in c_1 ; Risikoverbesserung wenn c_1 statt c_0 genutzt nur $O(1/n)$

ad (Q3/4) Pfanagl gilt klassen- (Hampel-Typ) aber nicht elementweise.

ad (Q5) auch im nicht-symmetr. Fall 2.(/k.)-Ord. optimale IC bestimmbar; führt i.a. aus Hampel-Klasse heraus; dabei aber stets Stützung! \rightsquigarrow keine Evidenz

Kommentare und Beweisskizze

- ges. Beweis für Median nötig, da Cramér-Bedingung verletzt
- Cross-Checks: durch Simulationen & num. Auswertungen
- viel expliziter als in Sattelpkt.app., c.f. Field & Ronchetti [90]
- (ex post) Erklärung f. Approx.-qualität bei Fraiman et al [01]

Beweisskizze

- Bedingen nach Zahl und Wert der Kontamination

	$K < 2r\sqrt{n}$	$2r\sqrt{n} \leq K < n/2$	$K \geq n/2$
$ t \leq 3b^2 \log(n)/n$	(I)	(II)	ausgeschlossen
$3b^2 \log(n)/n < t \leq Cn^2$	(III)		
$ t > Cn^2$	(IV)		

- zeige Vernachlässigbarkeit von (II), (III), (IV)
- Edgeworth-Entwicklung auf (I) und MAPLE für Termverwaltung

Optimale Stützhöhen c und (numer.) exakter $\max \text{MSE}$

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

c_0 1.Ord.opt wie gezeigt (Rieder)
 c_1 2.Ord.opt: wie gezeigt (P.R.)
 c_2 3.Ord.: num. Optim. d. MSE unter Hampel-typ ICs
 c_{FZY} num. Optimization gem. Vorschlag von Fraiman et al.
 c_{ex} num. Optim. d. exakten MSE

Vergleich k -Schritt- und M -Schätzer

Spezialisierung: 1-dim. symmetrische Lokation
Höherer-Ordnungs-Vergleich des maxMSE
Empirische Ergebnisse: Simulationsdesign
Empirische Ergebnisse: Simulationsergebnisse

Spezialisierung: 1-dim. symmetrische Lokation

Proposition (R.[11])

Sei $\Lambda_\theta(-\cdot) = -\Lambda_\theta(\cdot)$

- ▶ $\tilde{\psi}_\theta$ (1.Ord)-MSE-optimale IC zu Radius r (mit Stutzhöhe \tilde{b}_θ)
- ▶ $\psi_\theta^{(b_\theta)} = A_\theta \Lambda_\theta \min\{1, \frac{b_\theta}{|A_\theta \Lambda_\theta|}\}$ für ein $0 < b_\theta < \tilde{b}_\theta$.
- ▶ Seien S_n, S'_n entspr. M - und 1-Schritt-Schätzer zu $\tilde{\psi}_\theta$, wobei $\theta_n^{(0)}$ selbst ALE mit IC $\psi_\theta^{(b_\theta)}$ sei

Dann $\max\text{MSE}(S'_n) = \max\text{MSE}(S_n) + o(n^{-1/2})$

Bemerkung

Keine allgemeine Aussage zu (Q2):

Für (1.Ord)-MSE-suboptimale Hampel-IC können beide Fälle auftreten, $\max\text{MSE}(S'_n) \lesssim \max\text{MSE}(S_n) + o(n^{-1/2})$

Höherer-Ordnungs-Vergleich des maxMSE

Die glm. Entwicklung des MSE erlaubt folgenden Vergleich

Satz (R.[11])

$\theta \mapsto \psi_\theta$ sei k mal diff'bar in $L_1(P_\theta)$.

S_n, S'_n seien entspr. M - und k -Schritt-Schätzer zu ψ_θ .

Startsch. $\theta_n^{(0)}$ zu S'_n sei glm. konsist.+int'bar wie gerade

Dann erlaubt maxMSE von S_n, S'_n Entwicklungen der Ordg. k und $\max\text{MSE}(S'_n) = \max\text{MSE}(S_n) + o(n^{-(k-1)/2})$

- ▶ $n^{1/3}$ -konsistente $\theta_n^{(0)}$ s wie LMS-Regressions-Schätzer zulässig
- ▶ Explizite Terme nur bis Ord. 3 \implies Anwendung Theoreme für $k = 3$
- ▶ nicht- L_1 -glatte ICs wie Hampel-ICs durch ad-hoc Methoden abgedeckt

Empirische Ergebnisse: Simulationsdesign

- ▶ ideales Modell: $\mathcal{P} = \mathcal{N}(\theta, 1)$ at $\theta = 0$
- ▶ $M = 10000$ Versuchswiederholungen; Stichprobenumfang: $n = 5, 10, 30, 50, 100$
- ▶ Kontaminationsradii: $r = 0.1, 0.25, 0.5, 1.0$
- ▶ Kontaminierende Verteilung: Dirac-Vtlg. im Punkt 100
- ▶ ICs vom Huber-typ zu Stutzhöhen $c = 0.5, 0.7, 1, 1.5, 2$
- ▶ Schätzer:
 - ▶ M -Schätzer und
 - ▶ 1-Schritt-Schätzer mit Stichprobenmedian als Startschätzer

Empirische Ergebnisse: Simulationsergebnisse I

Empirischer und asymptotischer maxMSE zu $n = 30$, $c = 0.5$

r r/\sqrt{n}	M/1Schr	Simulation		Asymptotik		
		$\overline{\max\text{MSE}}_n$	[unt; ob]	n^0	$n^{-1/2}$	n^{-1}
0.00	1Schr	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	1Schr	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	1Schr	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

$\overline{\max\text{MSE}}_n$: Mittelwert der emp. Risiken, unt/ob: emp. 95% Konfidenzintervall
Asymptotik gemäß Termen der vorangegangenen Entwicklungen:

$$A_0 [+ r n^{-1/2} A_1 (+ n^{-1} A_2)]$$

Empirische Ergebnisse: Simulationsergebnisse I

Zahl der Iterationen I_n , die der M-Schätzer bei $n = 30$ und $c = 0.5$, sowie bei $n = 50$ und $c = 2.0$ braucht

r	Iterationen			
	$n = 30$ und $c = 0.5$		$n = 50$ und $c = 2.0$	
	\bar{I}_n	[unt; ob]	\bar{I}_n	[unt; ob]
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]

⇒ Mehraufwand im Vgl. zu 1Schr nicht gerechtfertigt

Zusammenfassung: Antworten auf Fragen (Q1)-(Q5) I

- (Q1) Einflussgrößen und Geschwindigkeit der Konvergenz?
 (A1) bei $r > 0$ Konvergenz eine Ordnung langsamer, nächste Terme in Entwicklung explizit als
 Fkt($\sup |\psi|, r$, Momente von ψ und deren Ableitungen)
- (Q2) Welche Konstruktion soll man verwenden?
 (A2) hins. Genauigkeit keine generelle Antwort; M- und k -Schritt-Schätzer werden mit wachsendem k immer ähnlicher; wg. Aufwand $k = 1, 2, 3$ -Schätzer zu empfehlen
- (Q3) (1.Ord)-opt. ALEs für finites n mit feinerer Asy. verbesserbar?
 (A3) Ja; durch niedrigere Stutzhöhe; explizite Terme hierfür; Im allgemeinen verlässt man Klasse der Hampel-ICs.

Zusammenfassung: Antworten auf Fragen (Q1)-(Q5) II

- (Q4) (Wie) Überträgt sich Pfanzagls Schlagwort [klass. Kontext]:
 "First order optimality implies second order optimality"?
- (A4) Die Aussage gilt (im wesentlichen) für die Klasse der Hampel-ICs; allerdings ist das optimale Element unterschiedlich für 1. und 2. Ordnung.
- (Q5) Theoretische Evidenz für Hampels Vermutung:
 "Finitely-optimal IC should be smooth"
- (A5) Keine Evidenz. Es muss für jede Iterationsstufe gestutzt werden.

mit M. Kohl und, seit 2009, N. Horbenko:

- ▶ **distr**-Paketfamilie
 - ▶ Verteilungen als Datenstrukturen: **distr**
 - ▶ Funktionale für Verteilungen: **distrEx**
- ▶ Wahrscheinlichkeitsmodelle
 - ▶ (glatte) parametrische Modelle (als Datenstruktur): **distrMod**
 - ▶ Zufallsvariablen als Abbildungen (als Datenstruktur): **RandVar**
- ▶ **RobASt**-Paketfamilie
 - ▶ Datenstrukturen für den Ansatz schrumpfender Umgebungen **RobAStBase**
 - ▶ 1.Ord.-MSE-opt IC + andere opt.-rob. Verfahren **ROptEst**

Bibliography II

- Ibragimov I. (1967): The Chebyshev-Cramér asymptotic expansions. *Theor. Probab. Appl.*, 12: 454–469.
- Ibragimov I. and Linnik Y. (1971): *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing Company, Groningen. Edited by J.F.C. Kingman.
- 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.
- 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.
- Kohl M. (2005): *Numerical contributions to the asymptotic theory of robustness*. Dissertation, Universität Bayreuth, Bayreuth.
- Kohl M. and P.R. (2010): R package **distrMod**: Object-Oriented Implementation of Probability Models. *J. Statist. Softw.* 35(10), 1–27.
- Kohl M., P.R., Rieder H. (2010): Infinitesimally Robust Estimation in General Smoothly Parametrized Models. *Stat. Methods Appl.*, 19, 333–354.
- Le Cam L. (1986): *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer.
- Pfanzagl J. (1979): First order efficiency implies second order efficiency. In: *Contributions to statistics, Jaroslav Hajek Mem. Vol.*, p. 167–196.
- (1985): *Asymptotic expansions for general statistical models. With the assist. of W. Wefelmeyer*, Vol. 31 of *Lecture Notes in Statistics*. Springer-Verlag.
- Pfaff T. (1977): *Existenz und asymptotische Entwicklungen der Momente mehrdimensionaler maximum likelihood-Schätzer*, Dissertation, Universität zu Köln, Köln.
- 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>
- Rieder H. (1980): Estimates derived from robust tests. *Ann. Stat.*, 8: 106–115.
- (1994): *Robust asymptotic statistics*. Springer Series in Statistics. Springer.

Bibliography I

- Barndorff-Nielsen O. and Cox D. (1994): *Inference and asymptotics*, Vol. 52 of *Monographs on Statistics and Applied Probability*. Chapman and Hall.
- Bhattacharya R. and Rao R. (1976): *Normal approximation and asymptotic expansions*. Wiley Series in Probability and Mathematical Statistics. John Wiley&Sons, Inc.
- Daniels H.E. (1954): Saddlepoint approximations in statistics. *Ann. Math. Statistics*, 25: 631–650.
- 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.
- Field C. and Ronchetti E. (1990): *Small sample asymptotics*, Vol. 13 of *IMS Lecture Notes - Monograph Series*. Institute of Mathematical Statistics, Hayward, CA.
- Fraiman R., Yohai V.J. and Zamar R.H. (2001): Optimal robust *M*-estimates of location. *Ann. Stat.*, 29(1): 194–223.
- 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.
- Hall P. (1992): *The bootstrap and Edgeworth expansion*. Springer Series in Statistics. Springer-Verlag.
- Hampel F.R. (1974): Some small sample asymptotics. In: *Proc. Prague Symp. Asympt. Stat., Vol. II, Prague 1973*, p. 109–126.
- Hoeffding W. (1963): Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.*, 58: 13–30.
- Huber P.J. (1968): Robust confidence limits. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 10: 269–278.
- (1981): *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley.
- (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 III

- Rieder H., Kohl M. and P.R. (2001): The Costs of not Knowing the Radius. *Stat. Methods and Appl.*, 17(1): 13–40.
- 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.
- Rousseeuw P.J. (1984): Least median of squares regression. *J. Am. Stat. Assoc.*, 79: 871–880.
- P.R. (2010a): Higher order asymptotics for the MSE of the sample median on shrinking neighborhoods. 25 Seiten. ArXiv 1006.0123.
- (2010b): Higher Order Expansion for the MSE of *M*-estimators on shrinking neighborhoods. 29 Seiten. ArXiv 1006.0037.
- (2010c): Consequences of Higher Order Asymptotics for the MSE of *M*-estimators on Neighborhoods. 25 Seiten. Eingereicht und in Revision. ArXiv 1006.0045.
- (2011): Higher order asymptotics for the MSE of *k*-step estimators on shrinking neighborhoods. In preparation.
- (2006): A Motivation for $1/\sqrt{n}$ -Shrinking-Neighborhoods. *Metrika*, 63(3): 295–307.
- P.R. and Kohl M. (2010a): General Purpose Convolution Algorithm for Distributions in *S*₄-Classes by means of FFT. 20 Seiten. ArXiv 1006.0764.
- (2010b): How to approximate finite sample risk of *M*-Estimators. 28 Seiten. <http://www.mathematik.uni-kl.de/~rueckdesc/pubs/howtoap.pdf>
- P.R. and Rieder, H. (2004): Optimal Influence Curves for General Loss Functions. *Statistics and Decisions*. 22: 201–223.
- P.R., Kohl M., Stabla T. and Camphausen F. (2006): *S*₄ Classes for Distributions. *R News* 6(2) 2–6.
- Yohai V.J. (1987): High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.*, 15(2): 642–656.