Computation of the Finite-Sample Risk of Robust Estimators via FFT

Matthias Kohl

UNIVERSITÄT BAYREUTH

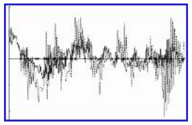
Mathematics VII

Workshop on Interval Probability 2004

LMU Munich

July 17th 2004

E-mail: matthias.kohl@uni-bayreuth.de



I Convolution via FFT

Assume two absolutely continuous distributions F_1, F_2 on $\mathbb{R}$ with unbounded support.

Step 1: Truncation

Step 2: Discretization on a real grid ($\rightarrow$ lattice distribution)

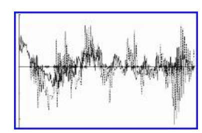
Step 3: Transformation to an integer grid ($\rightarrow$ integer latt. distr.)

Step 4: **Convolution via FFT** on integer grid (based on (circular) convolution theorem for discrete Fourier transforms)

Step 5: Back-transformation to real grid

Step 6: Smoothing (e.g. by linear functions)

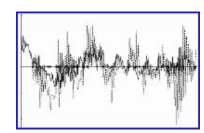
Step 7: Standardization



Implementation: R package `distr`, treats one-dimensional distributions by means of `S4` classes (on CRAN since April). [Ruckdeschel et al. (04)]

More details and examples: K. et al. (04)

Similar algorithms for the computation of compound distributions in insurance mathematics: Bertram (81), Grübel and Hermesmeier (99, 00) (including investigation of aliasing and discretization errors)



Example: Convolution of Exponential Distributions $\mathrm{Exp}\,(\lambda)$

n	ε	q	$d_v^\natural$	$d_\kappa^\natural$	comp. time
		10	2.1e−05	4.0e−05	0.1
2	**1e−08**	**13**	**3.3e−07**	**6.3e−07**	**0.8**
		15	3.2e−08	3.9e−08	3.5
		12	4.6e−06	4.6e−06	2.7
50	**1e−08**	**13**	**1.2e−06**	**1.0e−06**	**5.5**
		14	4.0e−07	3.2e−07	14.0

n: convolution power

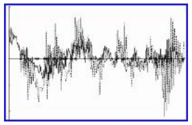
ε: truncation error (step 1)

2^q: number of lattice points (step 2)

$d_v^\natural$: (numerical) total variation distance of the densities

$d_\kappa^\natural$: (numerical) Kolmogorov distance of the cdfs.

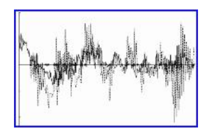
Matthias Kohl



Summary

- Results are **independent of the parameter λ**.
- Even **more accurate** in case of Normal distributions (also independent of the parameters).
- Algorithm is **numerically exact** for lattice distributions like $\mathrm{Binom}\,(n, p)$ or $\mathrm{Pois}\,(\lambda)$.
- Results can even be **improved under certain additional conditions** (e.g. smoothness of the densities $\Longrightarrow$ Richardson extrapolation) and **in certain special cases** (e.g. heavy-tailed distributions $\Longrightarrow$ exponential tilting).
  [Grübel and Hermesmeier (99, 00)]

UNIVERSITÄT BAYREUTH
Mathematics VII
Matthias Kohl



II One-dimensional normal location

II.1 Finite-sample solution

Contamination/total variation neighborhoods:

$$\mathcal{U}_{cv}(\theta) = \left\{ Q \in \mathcal{M}_1(\mathbb{B}) \,\middle|\, Q \geq (1-\varepsilon) P_\theta - \delta \right\}$$

where $\varepsilon, \delta \in [0,1)$ with $\varepsilon + \delta \in (0,1)$.

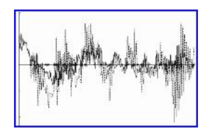
General assumptions:

- $P_\theta(du) = f(u-\theta)\lambda(du)$ with $\log f$ concave
- the observations are assumed to be i.i.d. (also under deviations Q_θ from P_θ); i.e., $Q_\theta^n = \bigotimes_{i=1}^n Q_\theta$ ($n \in \mathbb{N}$ fixed).

The finite-sample risk of an arbitrary estimator $S : \mathbb{R}^{2n} \to \mathbb{R}$ is

$$\mathrm{Risk}\,(S) = \sup_{\theta \in \mathbb{R}} \sup_{Q_\theta \in \mathcal{U}_{cv}(\theta)} \max \left\{ Q_\theta^n(S > \theta + \tau),\, Q_\theta^n(S < \theta - \tau) \right\}$$

for a given constant $\tau \in (0, \infty)$ (under-/overshoot probability).



Finite-sample solution for $P_\theta = N(\theta, 1)$:

Assume $\frac{\varepsilon+2\delta}{1-\varepsilon} < 2\Phi(\tau) - 1$. Then, the M estimator $\tilde{S}^{\mathrm{fi}}_{cv}$ satisfying

$$\sum_{i=1}^{n} \tilde{\psi}^{\mathrm{fi}}_{cv}(y_i - \tilde{S}^{\mathrm{fi}}_{cv}) = 0 \qquad \tilde{\psi}^{\mathrm{fi}}_{cv}(u) = u \min\Big\{1, \frac{b^{\mathrm{fi}}_{cv}}{|u|}\Big\}$$

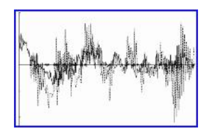
with equal randomization between the smallest and the largest solutions where b^{fi}_{cv} is the unique solution $b \in (0, \infty)$ to

$$\frac{\varepsilon+(1+\exp(-2b\tau))\delta}{1-\varepsilon} = \exp(-2b\tau)\Phi(\tau - b) - \Phi(-\tau - b)$$

is equivariant and minimax for all, arbitrary, estimators [Huber (68)].

One-dimensional simple linear regression: Rieder (89)

(assumptions: $F = f d\lambda$, $\log f$ concave, $K(x = 0) = 0$)



II.2 Asymptotic solution

Infinitesimal neighborhoods: Replace ε and δ by $\varepsilon_n := \varepsilon/\sqrt{n}$ and $\delta_n := \delta/\sqrt{n}$ where $\varepsilon, \delta \in [0, \infty)$. Moreover, $\tau \rightsquigarrow \tau_n := \tau/\sqrt{n}$.

General assumption: Finite Fisher-information of location
$\left[\mathcal{I}_\theta^{\mathrm{loc}} = \mathrm{E}_{P_\theta}(\Lambda_f^{\mathrm{loc}})^2 < \infty,\ \Lambda_f^{\mathrm{loc}}(u) = -f'/f(u)\right]$

The under-/overshoot probability of an asymptotically linear estimator $S = S_n$ with influence curve $\psi \in L_2(P_0)$ $(\mathrm{E}_{P_0}\,\psi = 0,$ $\mathrm{E}_{P_0}\,\psi\Lambda_f^{\mathrm{loc}} = 1)$ for given $\tau \in (0, \infty)$ is

$$\mathrm{asRisk}\,(S) = \Phi\left(\frac{\frac{\varepsilon+2\delta}{2}\left[\sup_{P_0}\psi - \inf_{P_0}\psi\right] - \tau}{\sqrt{\mathrm{E}_{P_0}\,\psi^2}}\right)$$

(by invariance: $\theta = 0$)

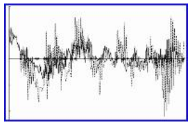
UNIVERSITÄT BAYREUTH

Mathematics VII

Matthias Kohl

Computation of the Finite-Sample Risk of Robust Estimators

07/17/2004



Asymptotic solution for $P_\theta = \mathcal{N}(\theta, 1)$:

Assume $(\varepsilon + 2\delta) < \sqrt{2/\pi}\,\tau$. Then,

$$\tilde\psi_{cv}^{\mathrm{as}}(u) = A_{cv}^{\mathrm{as}} u \min\left\{1, \frac{b_{cv}^{\mathrm{as}}}{|u|}\right\} \qquad (A_{cv}^{\mathrm{as}})^{-1} = 2\Phi(b_{cv}^{\mathrm{as}}) - 1$$

where b_{cv}^{as} is the unique solution $b \in (0, \infty)$ to

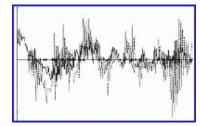
$$(\varepsilon + 2\delta) = 2\tau\big[\varphi(b) - b\Phi(-b)\big]$$

Construction of $\tilde S_{cv}^{\mathrm{as}}$ as M estimator with equal randomization between the smallest and largest solution.

[Rieder (94), Theorem 6.2.4]

One-dim. simple linear regression: Rieder (80, 81), K. (04)

(assumptions: $\mathcal{I}_\theta^{\mathrm{loc}} < \infty$, $\mathrm{E}\, x^2 < \infty$)



II.3 Computation of Finite-Sample Risk

We fix $n \in \mathbb{N}$ and consider $\tau_n \in (0, \infty)$ and radii $\varepsilon_n, \delta_n \in [0, 1)$.

M estimators: Given $b \in (0, \infty)$ we consider the M estimator S_{M} satisfying

$$\sum_{i=1}^{n} \psi(y_i - S_{\mathrm{M}}) = 0 \qquad \psi(u) = u \min\{1, b/|u|\}$$

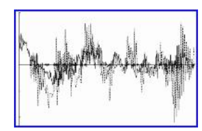
with **equal randomization between the smallest S' and the largest solutions S''**. S_{M} is equivariant and we have

$$\{S^{\prime[\prime]} > 0\} \subseteq \Big\{ \sum_{i=1}^{n} \psi(u_i) \underset{[\geq]}{>} 0 \Big\} \subseteq \{S^{\prime[\prime]} \geq 0\}$$

for any given $u_1, \ldots, u_n$. If Q is absolutely continuous we get

$$Q^n(S^{\prime[\prime]} > 0) = Q^n\Big(\sum_{i=1}^{n} \psi(u_i) \underset{[\geq]}{>} 0 \Big) = Q^n(S^{\prime[\prime]} \geq 0)$$

[Huber (81), Rieder (89)]



Attaining the finite-sample risk: [K. (04)]

The M estimator S_{M} attains the finite-sample risk under

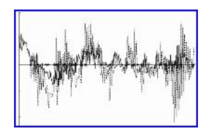
$$Q'_{-\tau_n} = (1-\varepsilon_n)\mathcal{N}\,(-\tau_n, 1) + \varepsilon_n H'_{-\tau_n}$$
$$Q''_{\tau_n} = (1-\varepsilon_n)\mathcal{N}\,(\tau_n, 1) + \varepsilon_n H''_{\tau_n}$$

in case of **contamination neighborhoods ($* = c$)** and under

$$Q'_{-\tau_n}(t) = \big(\Phi(t+\tau_n) - \delta_n\big)_+ + \delta_n H'_{-\tau_n}(t) \qquad (t \in \mathbb{R})$$
$$Q''_{\tau_n}(t) = \min\big\{\big[\Phi(t-\tau_n) + \delta_n H''_{\tau_n}(t)\big], 1\big\} \qquad (t \in \mathbb{R})$$

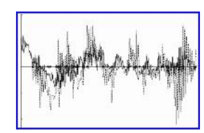
in case of **total variation neighborhoods ($* = v$)**
where $H'_{-\tau_n}$ and H''_{τ_n} are absolutely continuous and concentrated on $[\tau_n + b, \infty)$ and $(-\infty, -\tau_n - b]$, respectively.

One-dimensional simple linear regression: [K. (04)]



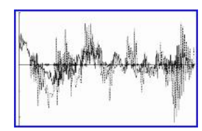
Algorithm A: This procedure **directly uses the distribution of ψ under $Q'_{-\tau_n}$, Q''_{τ_n}**. The n-fold convolution of this distribution is calculated using FFT.

Algorithm B: We determine the finite-sample risk by **splitting up** $\sum_{i=1}^{n} \psi(u_i)$ in $|\psi(u_i)| < b$ and $|\psi(u_i)| = b$; i.e., we obtain an **absolutely continuous** (conditional normal distributions) and a **discrete part** (random walk with values $\pm b$). The convolution of the discrete part can be done analytically, the convolution of the absolutely continuous part is done by FFT.



Similar Algorithms for the computation of the **finite-sample minimax MSE**; confer Ruckdeschel and K. (04).

Special Case ($n = 2$): If we choose ε ($* = c$), respectively, δ ($* = v$) such that $b_*^{\mathrm{fi}} = \tau_2$, respectively $b_*^{\mathrm{as}} = \tau_2$, we can calculate the **finite-sample risk analytically**.



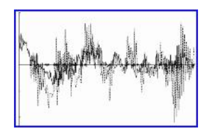
Precision of Algorithms A and B for $* = c$ and $n = 2$

ε	$\tau_2 = b_c^{\mathrm{as}}$	$\mathrm{Risk}(\tilde{S}_c^{\mathrm{as}})$	q	error_A	error_B
			10	5.1e−05	2.4e−08
0.0480	2.000	5.2560	**12**	**1.3e−05**	**1.5e−09**
			14	3.2e−06	9.5e−11
			10	8.4e−05	3.9e−08
0.2357	1.000	29.4290	**12**	**2.1e−05**	**2.4e−09**
			14	5.2e−06	1.5e−10

$\mathrm{Risk}(\tilde{S}_c^{\mathrm{as}})$: finite-sample risk of asymptotic minimax estimator.

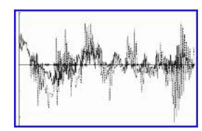
2^q: number of lattice points used in FFT algorithm.

$\mathrm{error}_{A/B}$: error of Algorithm A, resp. B.

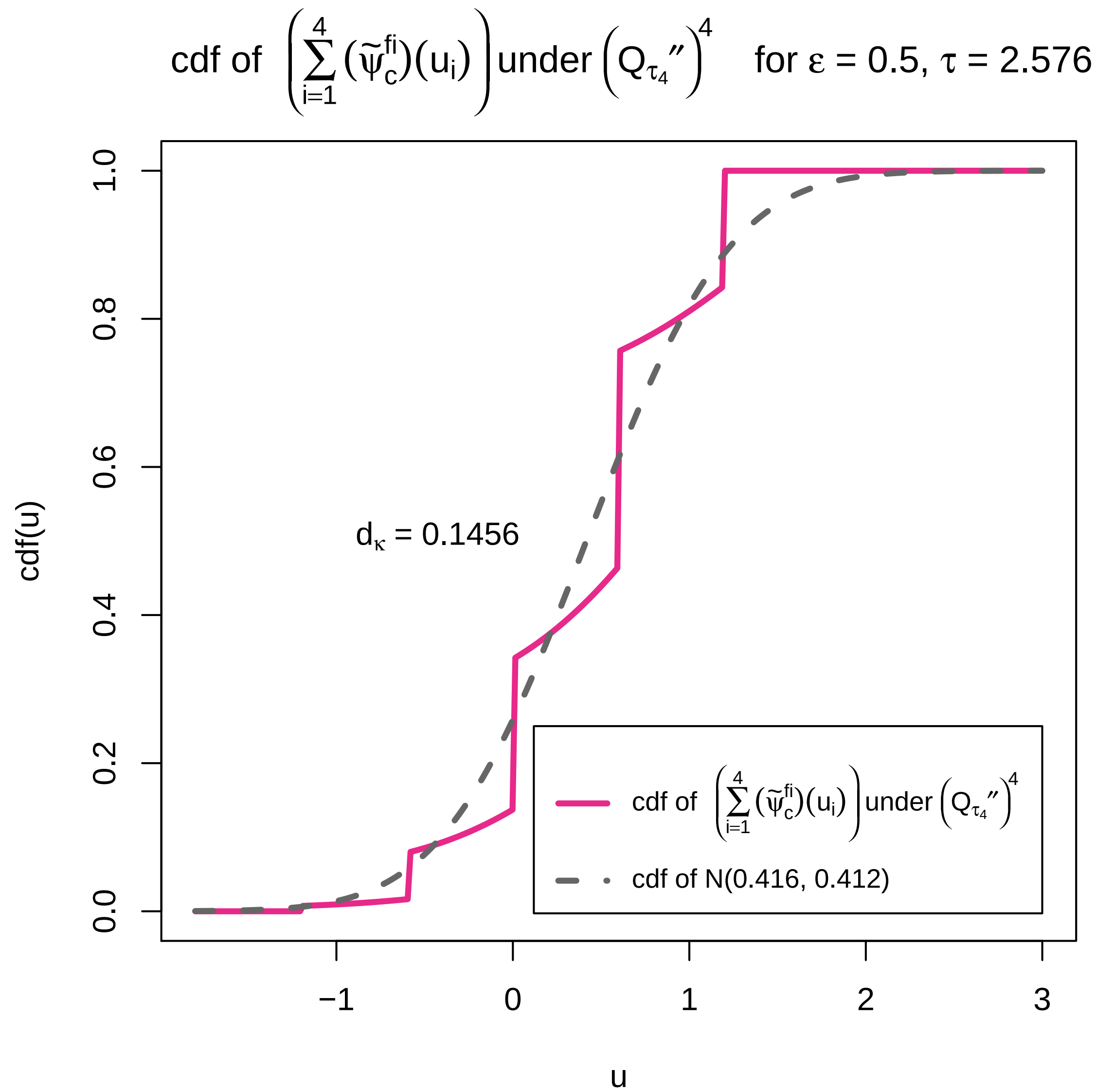


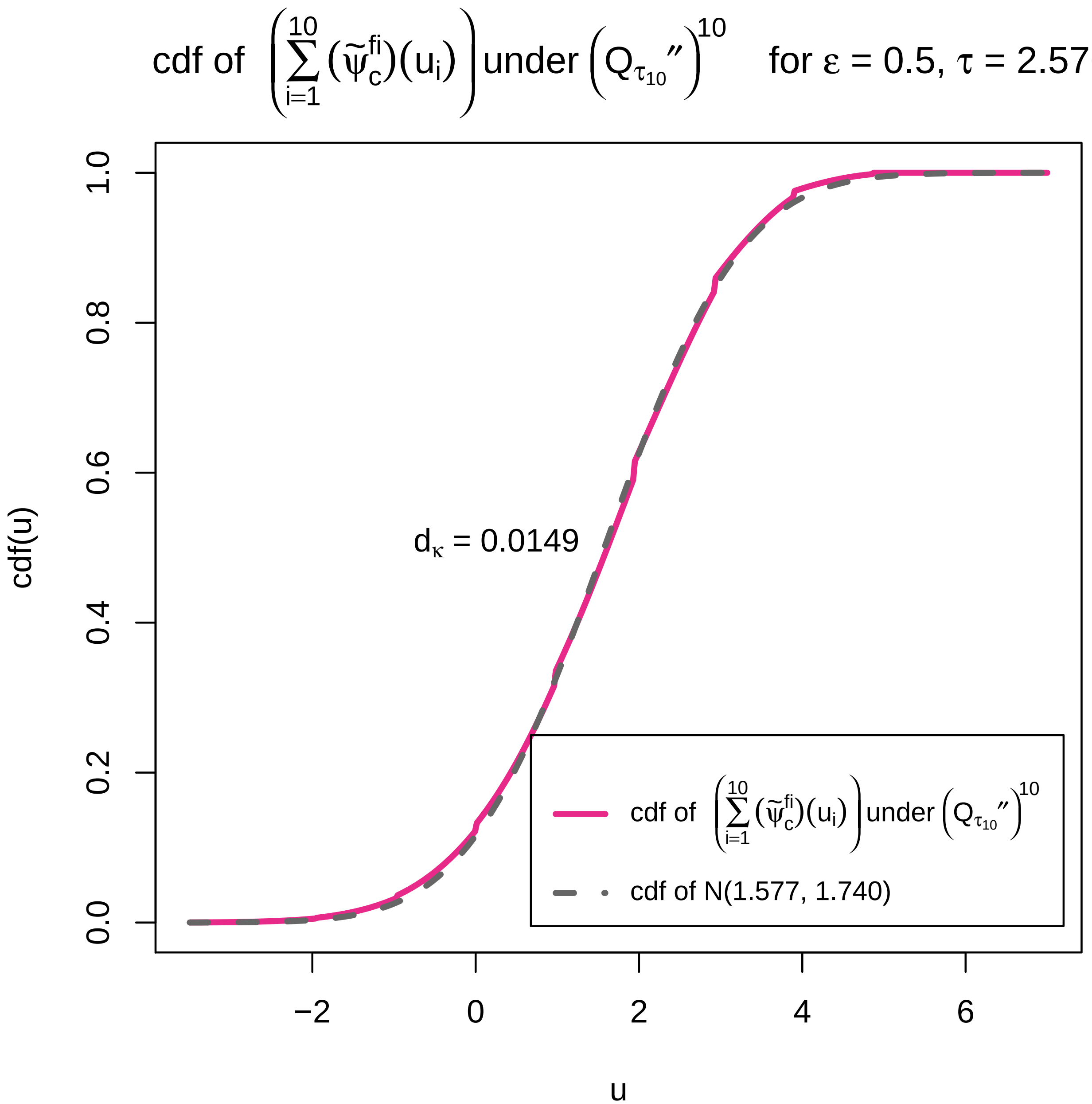
Summary

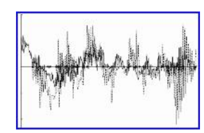
- The precision of the results is similar in case of
  - **total variation neighborhoods ($* = v$)**
  - **finite-sample minimax estimator $\tilde{S}_*^{\mathrm{fi}}$**
- Since the FFT convolution algorithm maintains its **high precision with increasing convolution power**, we expect the same behavior for Algorithms A and B. This is **confirmed by numerical simulations**.
- **Algorithm B is more accurate but Algorithm A is clearly faster**. In particular, in case of Algorithm B the computation time strongly increases with increasing sample size n.



II.4 Finite-sample distribution




cdf of $\left(\sum_{i=1}^{10}(\tilde{\psi}_c^{fi})(u_i)\right)$ under $\left(Q_{\tau_{10}}''\right)^{10}$ for $\varepsilon = 0.5$, $\tau = 2.576$
cdf(u)
0.0
0.2
0.4
0.6
0.8
1.0
$d_\kappa = 0.0149$
cdf of $\left(\sum_{i=1}^{10}(\tilde{\psi}_c^{fi})(u_i)\right)$ under $\left(Q_{\tau_{10}}''\right)^{10}$
cdf of N(1.577, 1.740)
−2
0
2
4
6
u



II.5 Contamination vs. Total Variation Neighborhoods

Optimal clipping bounds: [K. (04)] (by Taylor expansions)

$\boldsymbol{* = c}$: $b_c^{\rm fi} = b_c^{\rm as} + {\rm O}(n^{-1/2})$

$\boldsymbol{* = v}$: $b_v^{\rm fi} = b_v^{\rm as} + {\rm O}(n^{-1})$

(convergence from below)

Finite-sample risks: [K. (04)] (numerically)

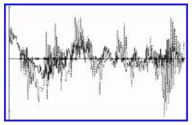
$\boldsymbol{* = c}$: ${\rm Risk}\,(\tilde S_c^{\rm fi}) = {\rm Risk}\,(\tilde S_c^{\rm as}) + {\rm O}(n^{-1/2})$

$\boldsymbol{* = v}$: ${\rm Risk}\,(\tilde S_v^{\rm fi}) = {\rm Risk}\,(\tilde S_v^{\rm as}) + {\rm O}(n^{-1})$

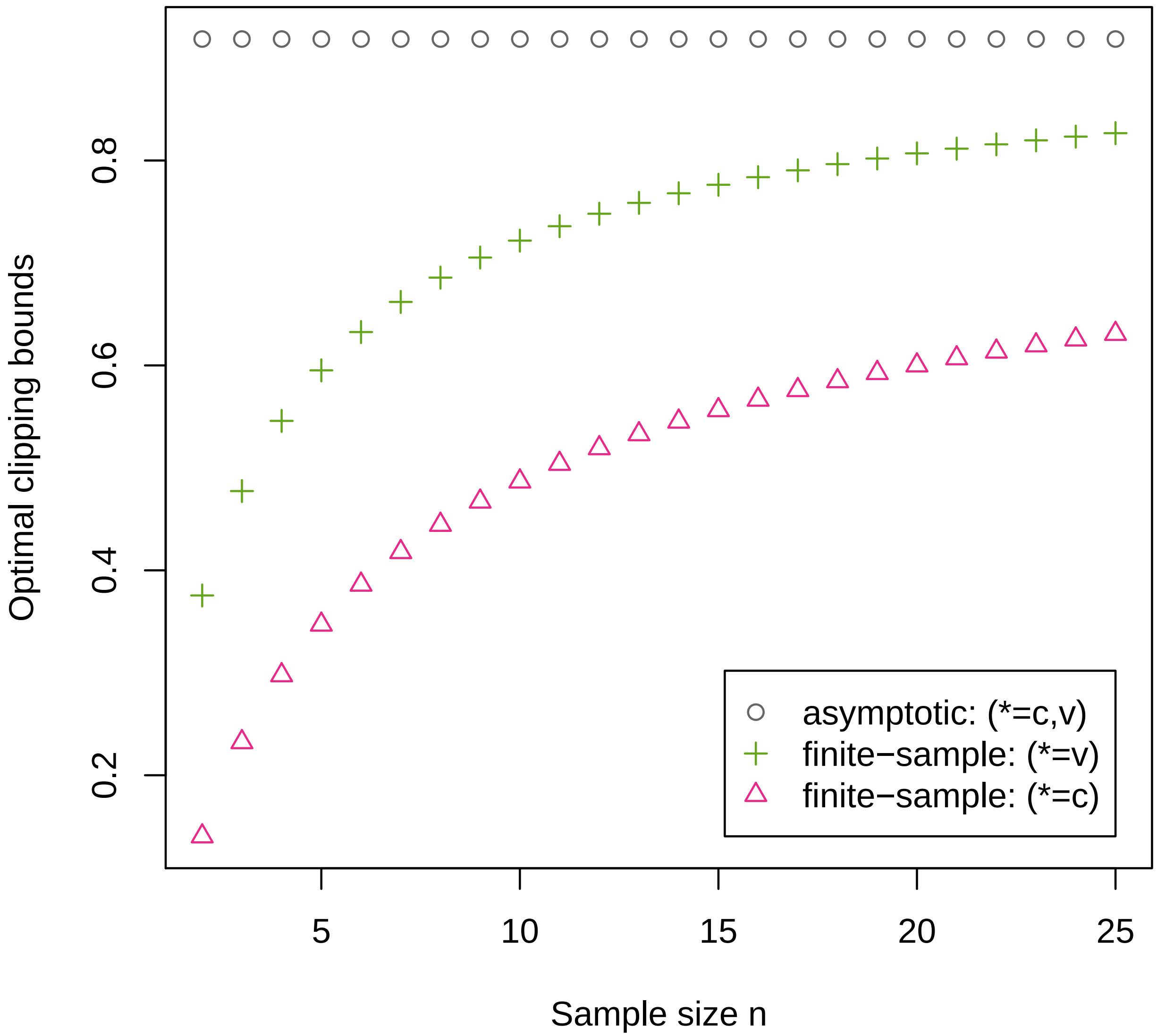
(convergence from above)

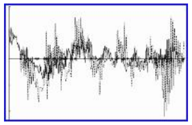
Simple linear regression: Essentially the same results. [K. (04)]

Relative risks: Various comparisons. [K. (04)]

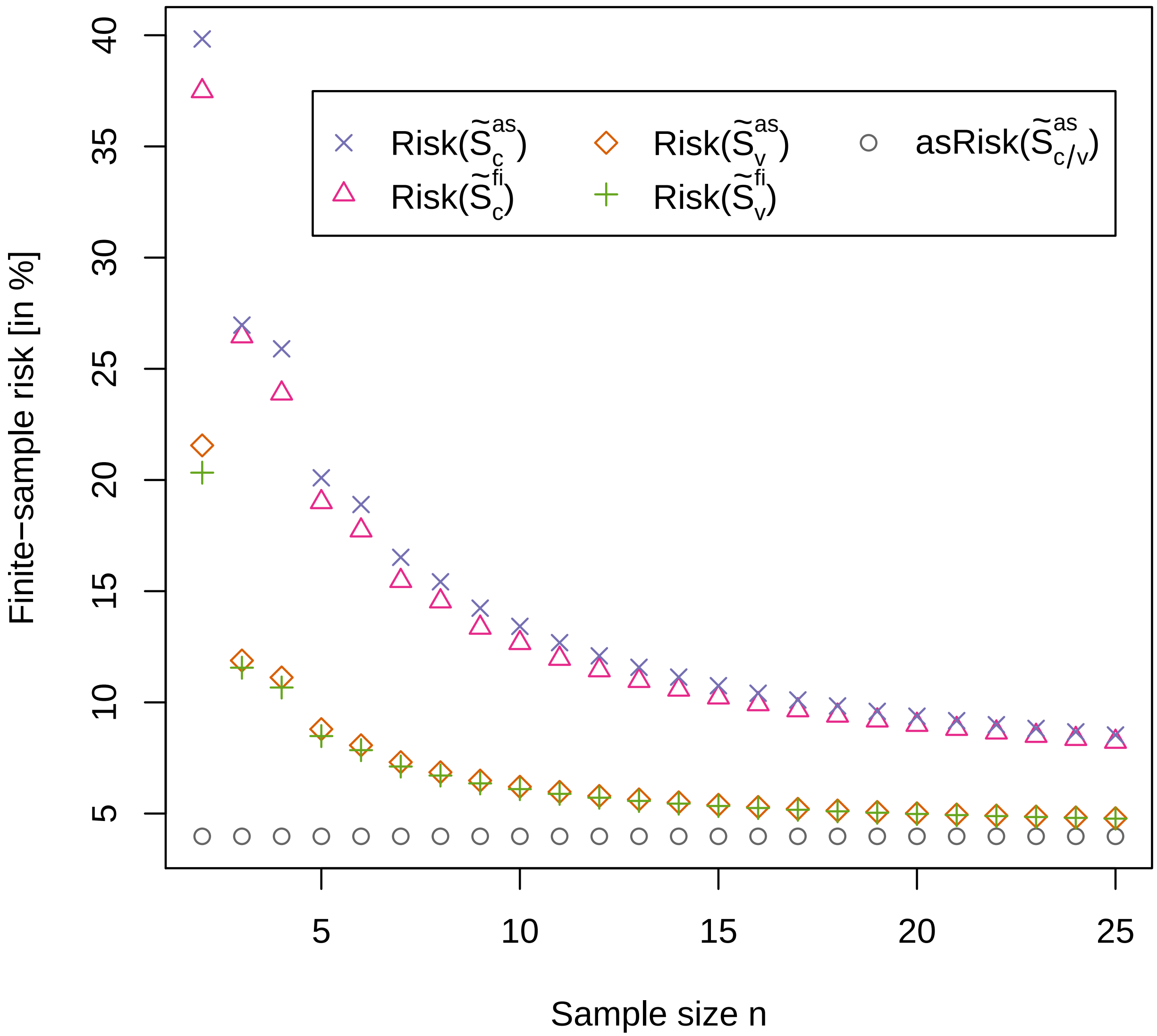


Optimal clipping bounds for ε = 0.5 = 2δ and τ = 2.576





Finite−sample and asymptotic risks for $\varepsilon = 0.5 = 2\delta$ and $\tau = 2.576$



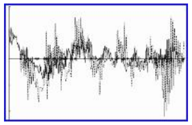
UNIVERSITÄT BAYREUTH

Mathematics VII

Matthias Kohl

Computation of the Finite-Sample Risk of Robust Estimators

07/17/2004



II.6 Higher Order Approximations

Edgeworth Expansion: The corresponding terms are

$$R^n\left(\sum_{i=1}^{n}\frac{\psi(u_i)-\mathrm{E}_R\,\psi}{\sqrt{\mathrm{Var}_R\,\psi}}<\sqrt{n}\,t\right)=\Phi(t)-\varphi(t)\left[\frac{\rho_R}{6\sqrt{n}}(t^2-1)\right.$$
$$\left.+\frac{1}{n}\left(\frac{\kappa_R}{24}(t^3-3t)+\frac{\rho_R^2}{72}(t^5-10t^3+15t)\right)\right]+\mathrm{O}(n^{-3/2})$$

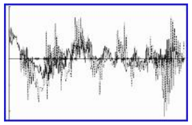
with $R = Q'_{-\tau_n}, Q''_{\tau_n}$ and

$$\rho_R=\mathrm{E}_R\left(\frac{\psi-\mathrm{E}_R\,\psi}{\sqrt{\mathrm{Var}_R\,\psi}}\right)^3 \qquad \text{and} \qquad \kappa_R=\mathrm{E}_R\left(\frac{\psi-\mathrm{E}_R\,\psi}{\sqrt{\mathrm{Var}_R\,\psi}}\right)^4-3$$

[Ibragimov (67), Field and Ronchetti (90)]

To determine an approximation of the finite-sample risk, choose

$$t=-\sqrt{n}\,\frac{\mathrm{E}_R\,\psi}{\sqrt{\mathrm{Var}_R\,\psi}}$$



Saddlepoint Approximations: An asymptotic expansion of the density of the M estimator S_M is

$$f_n(t) = \sqrt{\frac{n}{2\pi}}\, c^{-n}(t)\, \frac{A(t)}{\sigma(t)} \left[1 + \mathrm{O}(n^{-1})\right] \qquad t \in \mathbb{R}$$

with

$$c^{-1}(t) = \int \exp\{\alpha(t)\psi(u-t)\}\, r(u)\, du$$

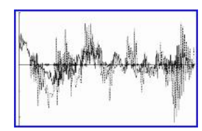
$$\sigma^2(t) = \int \psi(u-t)^2 c(t) \exp\{\alpha(t)\psi(u-t)\}\, r(u)\, du$$

$$A(t) = \int \left[\tfrac{\partial}{\partial t}\psi(u-t)\right] c(t) \exp\{\alpha(t)\psi(u-t)\}\, r(u)\, du$$

and $\alpha(t)$ is the solution $\alpha \in \mathbb{R}$ to

$$0 = \int \psi(u-t) \exp\{\alpha\psi(u-t)\}\, r(u)\, du$$

where $R = Q'_{-\tau_n}, Q''_{\tau_n}$ and $dR = r d\lambda$.[Field and Ronchetti (90)]



$* = c$: We fix $\varepsilon = 0.5$ and $\tau = \Phi^{-1}(0.995) \approx 2.576$ and let the sample size n increase ($\Longrightarrow$ actual sizes: ε_n, τ_n).

b_c^{as}	n	$\mathrm{Risk}^\natural(\tilde{S}_c^{\mathrm{as}})$	$\mathrm{Risk}_{\mathrm{E1}}$	$\mathrm{Risk}_{\mathrm{E2}}$	$\mathrm{Risk}_{\mathrm{S}}$
	2	39.83	34.92	36.50	40.17
	3	26.97	27.37	28.84	30.53
	4	25.90	22.88	24.01	24.76
	5	**20.10**	**19.90**	**20.75**	**21.08**
0.919	**10**	**13.42**	**13.14**	**13.41**	**13.40**
	100	5.73	5.73	5.73	5.73
	1000	**4.44**	**4.44**	**4.44**	**4.44**
	asymptotic risk: **3.98**				

$\mathrm{Risk}^\natural(\tilde{S}_c^{\mathrm{as}})$: finite-sample risk via Algorithm B $(n \le 10)$, resp. A $(n > 10)$.

$\mathrm{Risk}_{\mathrm{E}1/2}$ approximation via Edgeworth expansions up to first/second order.

$\mathrm{Risk}_{\mathrm{S}}$: approximation via saddlepoint approximations.

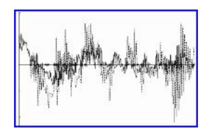
UNIVERSITÄT BAYREUTH

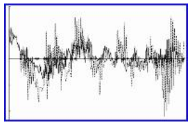
Mathematics VII

Matthias Kohl

Computation of the Finite-Sample Risk of Robust Estimators

07/17/2004



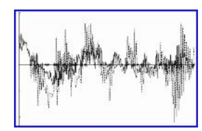


$* = v$: We fix $\delta = \varepsilon/2 = 0.25$ and $\tau = \Phi^{-1}(0.995) \approx 2.576$ and let the sample size n increase ($\Longrightarrow$ actual sizes: δ_n, τ_n).

b_v^{as}	n	$\mathrm{Risk}^{\natural}(\tilde{S}_v^{\mathrm{as}})$	$\mathrm{Risk}_{\mathrm{E1}}$	$\mathrm{Risk}_{\mathrm{E2}}$	$\mathrm{Risk}_{\mathrm{S}}$
0.919	2	21.56	16.14	18.04	19.13
	3	11.88	12.35	12.79	13.29
	4	11.12	10.18	10.24	10.48
	5	**8.80**	**8.83**	**8.80**	**8.92**
	10	**6.20**	**6.20**	**6.16**	**6.18**
	100	4.17	4.17	4.17	4.17
	1000	**4.00**	**4.00**	**4.00**	**4.00**
	asymptotic risk: **3.98**				

Summary

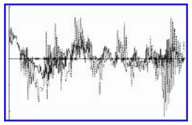
- **Convergence towards asymptotic risk** faster in case of total variation neighborhoods.
- Edgeworth expansion up to second order and saddlepoint approximation yield **comparable results**. In both cases we need about **$5 - 10$ observations** to get good approximations.
- For **smaller radii ε, δ** the Edgeworth expansion up to second order as well as the saddlepoint approximation yield **acceptable approximations down to sample size $n = 4$ or even $n = 3$**!
- The higher order approximations are **fast to compute** and the **computation time is independent of n** (in contrast to our Algorithms A and B).



III Conclusion

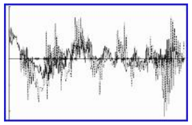
- We provide a way to compute a **very accurate numerical approximation of the finite-sample risk** and the finite-sample distribution of differently robust estimators using FFT. Thus, we are able to **check the asymptotics against finite-sample results**.
- The **speed of convergence towards the asymptotic values** is of order $\mathrm{O}(n^{-1/2})$ in case of contamination neighborhoods whereas it is of order $\mathrm{O}(n^{-1})$ in case of total variation neighborhoods.
- **Edgeworth expansions up to second order** as well as **saddlepoint approximations** yield good approximations of the finite-sample risk for sample sizes down to about 5 in this setup.

Matthias Kohl



References

Matthias Kohl



References

Bertram J. (1981): Numerische Berechnung von Gesamtschadenverteilungen. *Bl., Dtsch. Ges. Versicherungsmath.*, **15**: 175–194.

Field C. and Ronchetti E. (1990): *Small sample asymptotics.*. IMS Lecture Notes - Monograph Series. 13. Hayward, CA: Institute of Mathematical Statistics. 151 p. .

Grübel R. and Hermesmeier R. (1999): Computation of compound distributions. I. Aliasing errors and exponential tilting. *Astin Bull.*, **29**(2): 197–214.

—— (2000): Computation of compound distributions. II. Discretization errors and Richardson extrapolation. *Astin Bull.*, **30**(2): 309–331.

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.

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

Kohl M. (2004): *Numerical contributions to the asymptotic theory of robustness*. Dissertation, Universität Bayreuth, Bayreuth.

Kohl M., Ruckdeschel P. and Stabla T. (2004): General Purpose Convolution Algorithm for Distributions in S4 Classes by means of FFT. unpublished manuscript.

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

—— (1981): On local asymptotic minimaxity and admissibility in robust estimation. *Ann. Stat.*, **9**: 266–277.

—— (1989): A finite-sample minimax regression estimator. *Statistics*, **20**(2): 211–221.

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

Ruckdeschel P. and Kohl M. (2004): How to approximate the finite sample risk of M-estimators. unpublished manuscript.

Ruckdeschel P., Kohl M., Stabla T. and Camphausen F. (2004): S4 Classes for Distributions. http://www.uni-bayreuth.de/departments/math/org/mathe7/DISTR.