UNIVERSITÄT BAYREUTH

Lehrstuhl
Mathematik VII

S4-classes for infinitesimal robustness

International Workshop on Robust Statistics and R
Treviso

Oct 27th 2005

Peter Ruckdeschel
Matthias Kohl

E-mail: peter.ruckdeschel@uni-bayreuth.de
matthias.kohl@stamats.de

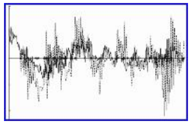
1 Organization

1 (a) (Co-)Authors

- Matthias Kohl: `matthias.kohl@stamats.de`
- [P.R.: `peter.ruckdeschel@uni-bayreuth.de`]
- Thomas Stabla: `statho3@web.de`
- Florian Camphausen: `fcampi@gmx.de`

1 (b) Organization in packages

- Packages for distributions
  - `distr`: basic package
  - `distrEx`: provides extensions to `distr` (e.g. expectation...)
  - `distrSim`: provides classes for simulations
  - `distrTEst`: provides classes for estmation and testing
  - `RandVar`: provides classes for random variables (also for dim>1)



- Packages for Robust Asymptotic Statistics
  - ROptEst: provides classes for optimally-robust estimation in infinitesimal robustness setup
  - RobLox*: optimally robust ICs for location and scale
  - RobRex*: optimally robust ICs for regression and scale
  - ROptRegTS*: infinitesimal robustness for regression and time series models

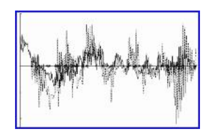
  *: not here

1 (c) Availability

- distr, distrEx, distrSim, distrTEst, RandVar: available on CRAN;
- ROptEst, RobLox, RobRex, ROptRegTS available on CRAN soon; preliminary versions under http://stamats.de/RobASt.htm

1 (d) References

- http://www.uni-bayreuth.de/departments/math/org/mathe7/
  /RUCKDESCHEL/pubs/distr.pdf
- http://stamats.de/ThesisMKohl.pdf

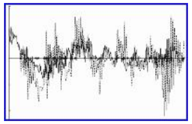


2 Packages for distributions

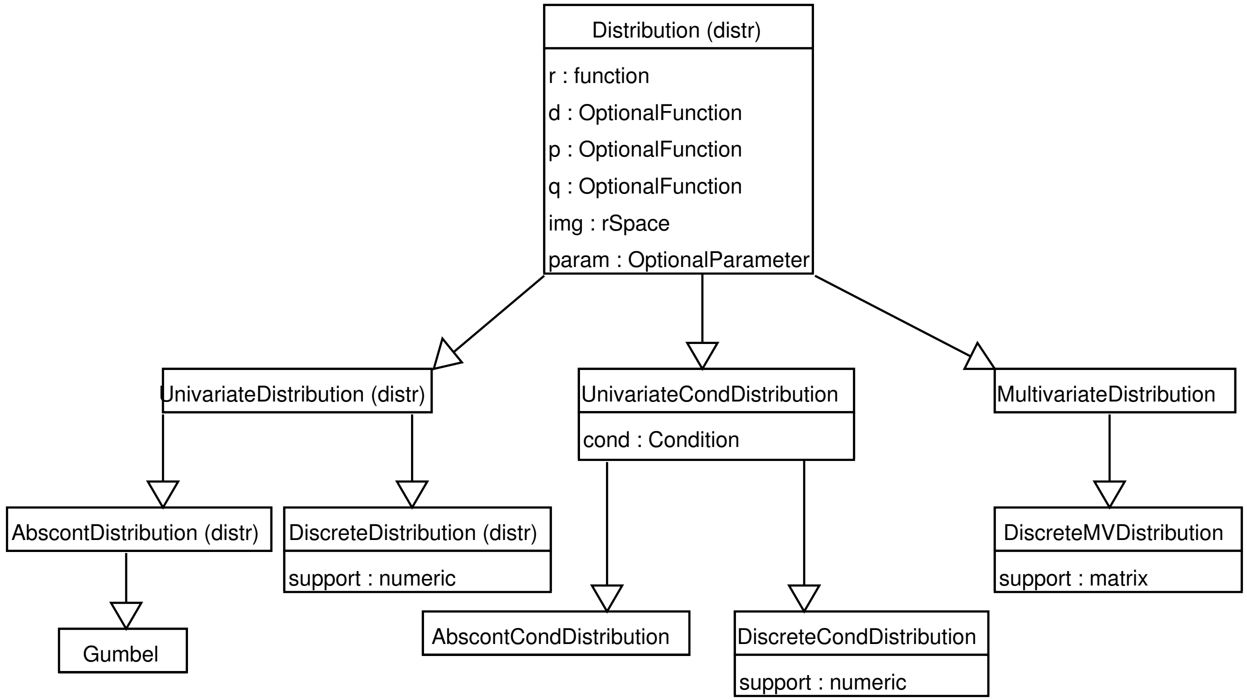
2 (a) Basic package: `distr`

Motivation

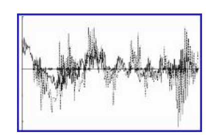
- realization of distributions in `S`:
  - highly-skilled implementations for virtually any useful (univariate) distribution
  - naming convention: `[prefix]<name>`
    `[prefix]` $\hat{=}$ `r`, `d`, `p`, or `q` and `<name>` is the name of the distribution, e.g. `norm`
- limitation: how to formulate an algorithm once for all distributions?
- works, but. . . : **eval**(**parse** (....))
- possible with `distr`: — see `R`-example



Organization in classes

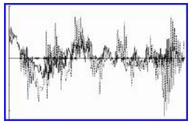


- further subclasses of `AbscontDistribtion`: `Beta`, `Cauchy`, `Chisq`, `Exp`, `Fd`, `Gammad`, `Logis`, `Lnorm`, `Norm`, `Td`, `Unif`, `Weibull` (all from base package)
- further subclasses of `DiscreteDistribtion`: `Binom`, `Dirac`, `Geom`, `Hyper`, `Nbinom`, `Pois` (all from base package)



Methods

- overloaded: operators "+", "-", "*", "/" — e.g. Y←(3∗X+5)/4
- group `math` of unary mathematical operations is available for objects of class `Distribution` e.g. **exp**(**sin**(3∗X+5)/4)
- `RtoDPQ`: default method for filling slots `d`, `p`, `q` on basis of simulations
- a default convolution method for two independent r.v.'s by means of FFT; c.f. K., R., & Stabla[04]
- particular methods for **plot**, **summary**,. . .



2 (b) `distrEx`

Functionals for distributions

for a distribution `F` on $\mathbb{B}^k$, e.g. D1 ← **Norm**(**mean**=2)

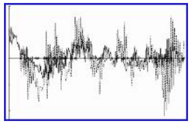
- expectation E of a distribution (with or without transformation); see `R`-example
- for robust statistics: truncated moments

$$\begin{aligned} \texttt{m1df(F,t)} &:= \int \mathrm{I}_{(-\infty;t]}(x)x\,F(dx), \\ \texttt{m2df(F,t)} &:= \int \mathrm{I}_{(-\infty;t]}(x)x^2\,F(dx) \end{aligned}$$

- easy but yet to be done: variance, median, MAD, IQR, ...

Distances between distributions

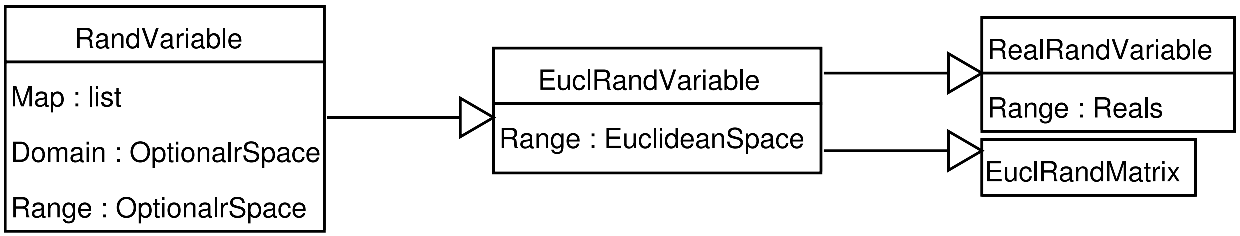
- `ContaminationSize` —pseudo-distance: $d_c(P,Q) := \inf\{r > 0 \mid \exists \text{ p.m. } H : Q = (1-r)P + rH\}$
- Hellinger-, Kolmogorov-, total-variation-distance
- compare `R`-example



2 (c) `RandVar`

Random variable as a class concept

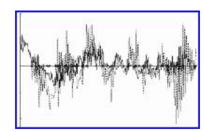
- Definition



Mathematical operations

- there are **many**. . .
- for `"RealRandVector"`, `"EuclRandVector"` and `"EuclRandMatrix"` all arithmetic und matrix operations are available and are used as with numeric vectors and matrices
- also: group `math` may be applied to it
- e.g. for random variable objects `X` and `Y`, a numerical vector `v` and a matrix `M` (with compatible dimensions), we can generate `exp(X - v)`, `X %*% Y` or `M %*% Y` where "%*%" stands for matrix multiplication.

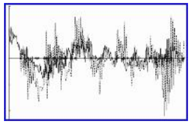
Peter Ruckdeschel
Matthias Kohl



3 Skipped in the talk:
Infinitesimal Robustness in 11 slides

3 (a) References:

- P.J. Bickel (1981): Quelques aspects de la statistique robuste. In *Ecole d'été der probabilités de Saint Flour IX-1979*, Lect. Notes Math. 876, p.2–72.
- H. Rieder (1994): *Robust asymptotic statistics*. Springer.
- ——, M.K., and P.R. (2001): The Costs of not Knowing the Radius. Submitted. Appeared as discussion paper Nr. 81. SFB 373 (Quantification and Simulation of Economic Processes), Humboldt University, Berlin; `http://www.uni-bayreuth.de/departments/math/org/mathe7/` `/RIEDER/pubs/RR.pdf`.
- P.R. and H. Rieder (2004) Optimal IC's for general loss functions. *Statistics and Decisions* **22**, p.201–223
- M.K.(2005) *Numerical contributions to the asymptotic theory of robustness*. Dissertation, Universität Bayreuth, Bayreuth. Available under `http://stamats.de/ThesisMKohl.pdf`



3 (b) L_2-differentiable model

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

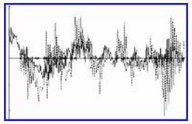
- Examples:
  - Gaussian location: $\mathcal{P}_1 = \{\mathcal{N}(\theta, 1) \mid \theta \in \Theta\}$, $\Theta = \mathbb{R}$
  - Gaussian scale: $\mathcal{P}_2 = \{\mathcal{N}(1, \theta(= \sigma^2)) \mid \theta \in \Theta\}$, $\Theta = (0, \infty)$
  - Gaussian location and scale: $\mathcal{P}_3 = \{\mathcal{N}(\theta_1, \theta_2) \mid \theta \in \Theta\}$, $\Theta = \mathbb{R} \times (0, \infty)$
- L_2-differentiability
  - is a smoothness assumption

    :: $\sqrt{dP_{\theta+h}} = \sqrt{dP_\theta}\,(1 + \frac{1}{2}\Lambda_\theta^\tau h) + \mathrm{o}(|h|)$

    i.e., for any sequence h_n in $\mathbb{R}^k$, $h_n \to 0$, $n \to \infty$ and any prob. measure $\mu \gg P_\theta, P_{\theta+h_n}$ $\forall n$

$$\int \Big(\sqrt{p_{\theta+h_n}} - \sqrt{p_\theta}\,(1 + \tfrac{1}{2}\Lambda_\theta^\tau h_n)\Big)^2 \, d\mu = \mathrm{o}(|h_n|^2)$$

    for the μ-densities $p_\theta, p_{\theta+h_n}$ of $P_\theta, P_{\theta+h_n}$.
  - also required: Fisher-information $\mathcal{I}_\theta := \int \Lambda_\theta \Lambda_\theta^\tau \, dP_\theta$ is finite



- Consequence:
  - $P^n_{\theta+h/\sqrt{n}}$ and P^n_θ are contiguous
  - Loglikelihood-expansion:

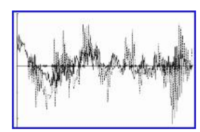
    $$\log dP^n_{\theta+h/\sqrt{n}}/P^n_\theta = \tfrac{1}{\sqrt{n}}\sum_i h^\tau \Lambda_\theta(x_i) - \tfrac{1}{2}h^\tau \mathcal{I}_\theta h + \mathrm{o}_{P^n_\theta}(1)$$

$\Longrightarrow$ model is LAN (locally asymptotically normal)

3 (c) differentiable parameter transformation

$\tau\colon \mathbb{R}^k \longrightarrow \mathbb{R}^p$, $\tau'(\theta) = D = D(\theta)$

- Examples:
  - estimation of sdv: in $\mathcal{P}_2$, we want to estimate $\sqrt{\theta}$, hence $\tau(x) = \sqrt{x}$
  - estimation/testing in the presence of a nuisance parameter: in $\mathcal{P}_3$, we want to estimate θ_1, but don't know θ_2

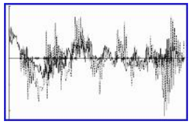


3 (d) Influence curves and asymptotically linear estimators

Influence curve (IC)

$$\eta_\theta \in L_2^p(P_\theta) \quad \text{s.t.} \quad \mathrm{E}_\theta\, \eta_\theta = 0, \quad \mathrm{E}_\theta\, \eta_\theta \Lambda_\theta^\tau = D \qquad (\mathrm{E}_\theta = \mathrm{E}_{P_\theta})$$

- here: IC not as a (functional) derivative but as a possible linearization of an estimator
- conditions guarantee / are equivalent to local uniform asymptotic normality of a corresponding ALE



Asymptotically linear estimators (ALEs)

estimators with expansion

$$\sqrt{n}\,(S_n-\theta)=\tfrac{1}{\sqrt{n}}\sum_{i=1}^n \eta_\theta(X_i)+\mathrm{o}_{P_\theta^n}(n^0)$$

for some IC η_θ

- Let S_n be an ALE with IC η_θ

THEN for any $0<C<\infty$, d_π the Prokhoroff distance and $n\to\infty$,

$$\sup_{|h|\le C} d_\pi\Big(\sqrt{n}\,(S_n-\theta)\circ P^n_{\theta+h/\sqrt{n}}\,,\mathcal{N}_p(h,\int \eta_\theta\eta_\theta^\tau\,dP_\theta)\Big)\to 0$$

- hence: S_n is Hájek regular

- Examples
  - in $\mathcal{P}_1$: $S_n=\bar X_n$ — $\eta_\theta(x)=x-\theta$,
  - in $\mathcal{P}_1$: $S_n=\mathrm{Median}_n$ — $\eta_\theta(x)=\sqrt{\pi/2}\,\mathrm{sign}(x-\theta)$

UNIVERSITÄT BAYREUTH

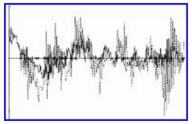
Mathematics VII

Peter Ruckdeschel

Matthias Kohl

S4-classes for infinitesimal robustness

10/27/2005



3 (e) One-step-estimators

defined to starting estimate θ_0 and IC η as

$$S_n^{(1)} := \tau(\theta_0) + \frac{1}{n}\sum_{i=1}^{n} \eta_{\theta_0}(X_i)$$

Assumptions:

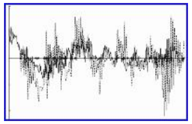
- θ_0 must be $\sqrt{n}$-consistent:
  $\sqrt{n}(\theta_0 - \theta) = \mathrm{O}_{Q_n^{(n)}}(1)$ uniformly for all $Q_n^{(n)}$ in the neighborhood(see below)
- IC η_θ is bounded and, uniformly over all arg's x, continuous in θ:

$$\lim_{h\to 0} \sup_x |\eta_{\theta+h}(x) - \eta_\theta(x)| = 0$$

THEN $S_n^{(1)}$ is an ALE:

$$S_n^{(1)} - \tau(\theta) = \tfrac{1}{n}\sum_{i=1}^{n} \eta_\theta(X_i) + \mathrm{o}_{P_\theta^n}(n^{-1/2})$$

UNIVERSITÄT BAYREUTH
Mathematics VII
Peter Ruckdeschel
Matthias Kohl



3 (f) (Shrinking) neighborhood system $\mathcal{U}_*(P_\theta, r)$ to radius r

- To P_θ^n consider $Q_n^{(n)} = \bigotimes_{i=1}^n Q_{n,i}$ for

$* =$c convex contaminations: $Q_{n,i} = (1 - r/\sqrt{n}\,)P_\theta + r/\sqrt{n}\,H_{n,i}$ for arbitrary, unknown $H_{n,i}$

$* =$v total variation: $Q_{n,i}$ such that $d_v(Q_{n,i}, P_\theta) \le r/\sqrt{n}$ for $2d_v(P,Q) = \int |dP - dQ|$

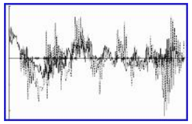
$* =$h Hellinger: $Q_{n,i}$ such that $d_h(Q_{n,i}, P_\theta) \le r/\sqrt{n}$ for $2d_h(P,Q)^2 = \int (\sqrt{dP} - \sqrt{dQ})^2$

THEN for all such $Q_n^{(n)} \in \mathcal{U}_*(P_\theta, r)$

$$\sqrt{n}\,\Big(S_n^{(1)} - \tau(\theta) - \tfrac{1}{n}\textstyle\sum_{i=1}^n \int \eta_\theta\, dQ_{n,i}\Big) \circ Q_n^{(n)} \xrightarrow{\mathrm{w}} \mathcal{N}_p(0, \mathrm{E}_\theta\, \eta_\theta \eta_\theta^\tau)$$

Peter Ruckdeschel

Matthias Kohl



3 (g) Optimally robust estimators

Maximal bias: For $\mathcal{U}_*(P_\theta, r)$

- $r\omega_*(\eta_\theta) := \sup_{Q_n^{(n)}} \frac{1}{n}\sum_{i=1}^n \int \eta_\theta \, dQ_{n,i}$

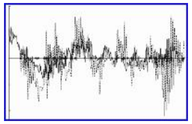
THEN

$* =$c $\omega_c(\eta_\theta) = \sup|\eta_\theta|$

$* =$v $\omega_v(\eta_\theta) = \sup \eta_\theta - \inf \eta_\theta \qquad (p = 1)$

$* =$h $\omega_h(\eta_\theta) \doteq \sqrt{8}\;\text{maximal eigenvalue}(\mathrm{E}_\theta\, \eta_\theta \eta_\theta^\tau)$

Peter Ruckdeschel
Matthias Kohl



MSE-optimal IC

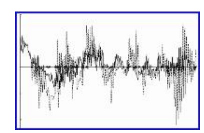
- maximal asymptotic MSE on $\mathcal{U}_*(P_\theta, r)$:
  $\mathrm{asMSE}(\eta, r) = \mathrm{E}_\theta\, |\eta_\theta|^2 + r^2 \omega_*^2(\eta_\theta)$
- MSE problem: To given r, find IC $\hat\eta_r$ minimizing asMSE
- solution: to given θ (suppressed in notation)

$* =$c Let $Y = A\Lambda - a$; then $\hat\eta_r = Y \min\{1, b/|Y|\}$ (Hampel-form) for $A \in \mathbb{R}^{p \times k}$, $a \in \mathbb{R}^p$ and $b > 0$ such that η is an IC and $r^2 b = \mathrm{E}(|Y| - b)_+ =: \gamma_c$

$* =$v $\hat\eta_r = c \wedge A\Lambda \vee (c + b)$ for $A \in \mathbb{R}^{1 \times k}$, $c \in (-b, 0)$ and $b > 0$ such that η is an IC and $r^2 b = \mathrm{E}(c - A\Lambda)_+ =: \gamma_v$ $(p = 1)$

$* =$h $\hat\eta_r = D\mathcal{I}^{-1}\Lambda$

  – examples in the R-scripts



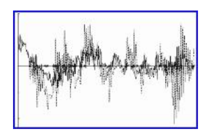
G-optimal IC

- fix θ; assume that maximal asymptotic risk on $\mathcal{U}_*(P, r)$ representable as
  $\tilde{G}(\eta, r) = G(r\omega_*(\eta), \sigma_\eta)$ for
  - $\sigma_\eta^2 = \mathrm{E}_P\,|\eta|^2$
  - $G = G(w, s)$ convex, isotone in both arguments, totally differentiable, bounded away from the minimum for $w \to \infty$

THEN for $* =$c or $* =$v($p = 1$) we obtain the same type of solutions as those with asMSE, but b determined as
$r\sigma_\eta G_w(rb, \sigma_\eta) = \gamma_* G_s(rb, \sigma_\eta)$

- examples:
  $G = \int |x|^q \, d\mathcal{N}(w, s)$ (L_q-risk),
  $G = \int \mathrm{I}(|x| > \tau) \, d\mathcal{N}(w, s)$ (Maximin covering probability)

Peter Ruckdeschel
Matthias Kohl



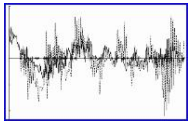
3 (h) Unknown radius r

- situation: r is not known, only available information $r \in [r_l, r_u]$
- relative inefficiency of η_r when used at radius s:
  $\rho(r,s) := \mathrm{asMSE}(\eta_r, s)/\mathrm{asMSE}(\eta_s, s)$
- maximal relative inefficiency of η_r: $\hat\rho(r) := \sup_{s\in[r_l,r_u]} \rho(r,s)$
- minimax radius/inefficiency: $r = r_0$ such that $\hat\rho(r)$ is minimal
- correspondingly with $\tilde G$ replacing asMSE

THEN for all homogeneous G (i.e. $G(\nu w, \nu s) = \nu^\alpha G(w,s)$),
η_{r_0} (parametrized by b) does *not* depend on G!

- examples for minimax radius in the R-scripts
  (see S4-method `radiusMinimaxIC`)

UNIVERSITÄT BAYREUTH
Mathematics VII
Peter Ruckdeschel
Matthias Kohl



4 Packages for Robust Asymptotic Statistics

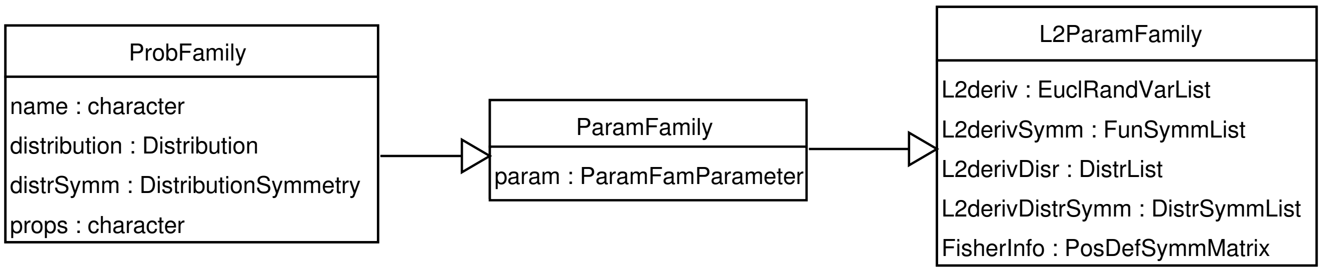
4 (a) "Main ingredients" of infinitesimal robustness

(more detailed information on the distributed slides)
compare Rieder[94], Rieder, K., R.[2K]

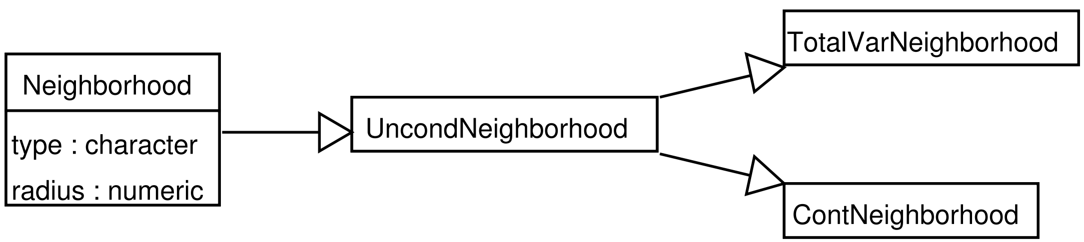
- L_2-differentiable model $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^k$ open
- differentiable parameter transformation $\tau\colon \mathbb{R}^k \to \mathbb{R}^p$, $\tau'(\theta) = D = D(\theta)$
- influence curves (IC): $\eta_\theta \in L_2^p(P_\theta)$ s.t. $\quad \mathrm{E}_{P_\theta}\, \eta_\theta = 0, \quad \mathrm{E}_{P_\theta}\, \eta_\theta \Lambda_\theta^\tau = D.$
- asymptotically linear estimators (ALEs): with expansion
  $$\sqrt{n}\,(S_n - \theta) = \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n \eta_\theta(X_i) + \mathrm{o}_{P_\theta^n}(n^0)$$
  for some IC η_θ
- one-step-estimators to starting estimate θ_0 and IC η
  $$S_n^{(1)} := \tau(\theta_0) + \frac{1}{n} \sum_{i=1}^n \eta_{\theta_0}(X_i)$$
- (shrinking) neighborhood system to radius r given or unknown
- risk

4 (b) Classes

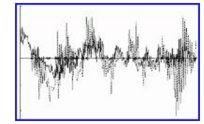
- L_2-differentiable model:



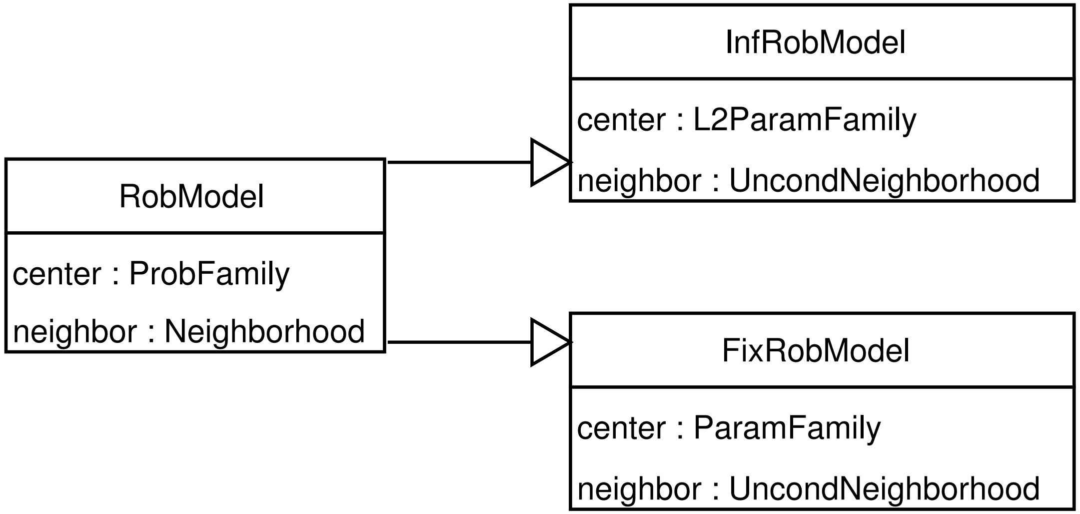
- neighborhood system to some given radius r



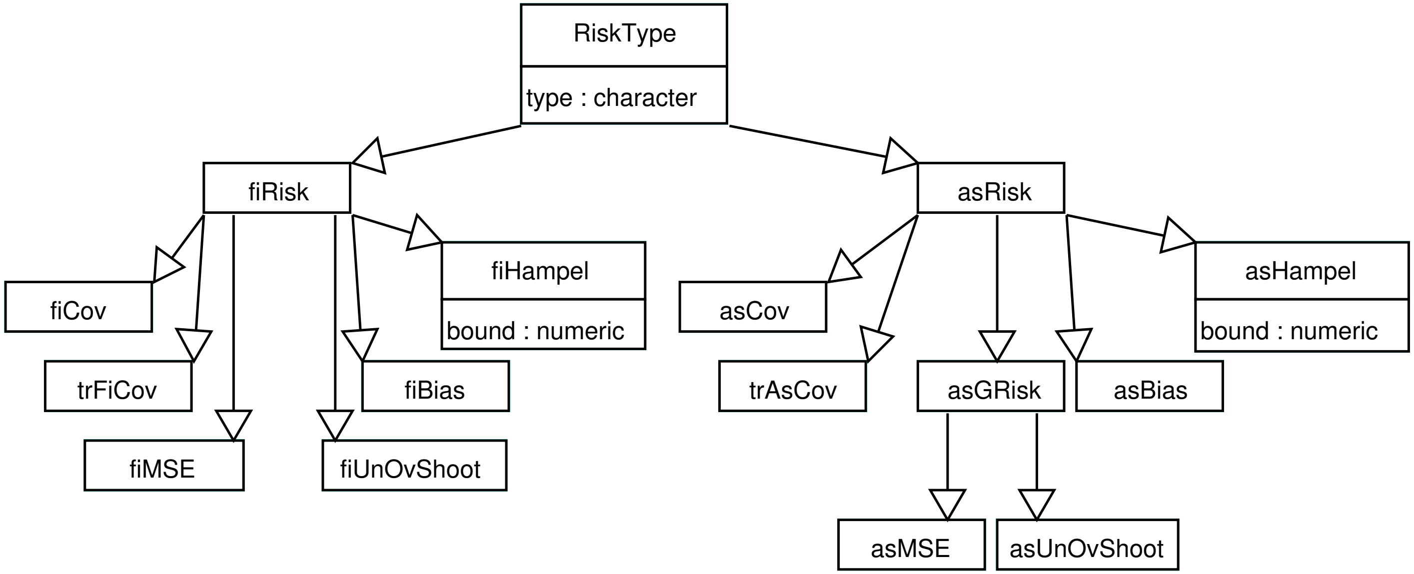
Peter Ruckdeschel
Matthias Kohl



- robust model

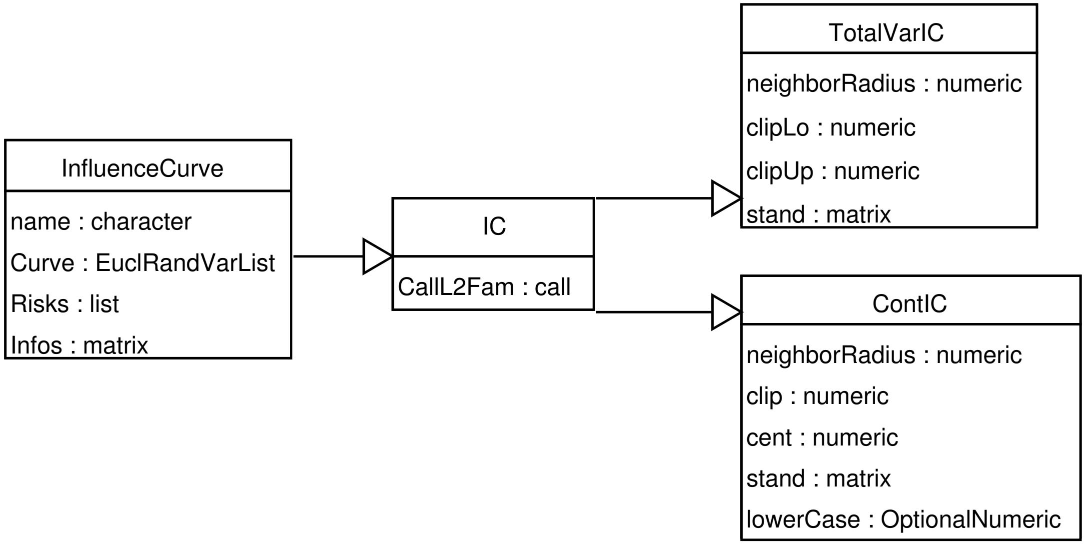


- risk



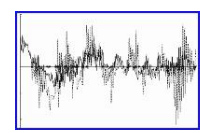
Peter Ruckdeschel
Matthias Kohl

- IC



4 (c) Methods

- accessor and replacement functions, **show**, **plot**
- `addInfo`, `addProp`, `addRisk`
- `checkL2deriv`, `checkIC`, `evalIC`, `getRiskIC`, `infoPlot`, `ksEstimator`, `leastFavorableRadius`, `locMEstimator`, `oneStepEstimator`, `optIC`, `optRisk`, `radiusMinimaxIC`
- easy generating functions for implemented L_2-families like `NormLocationScaleFamily`, `BinomFamily`



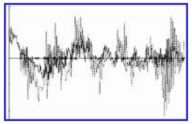
Special meta-information slots

- information gathered during generation of objects is stored in information slots, e.g.

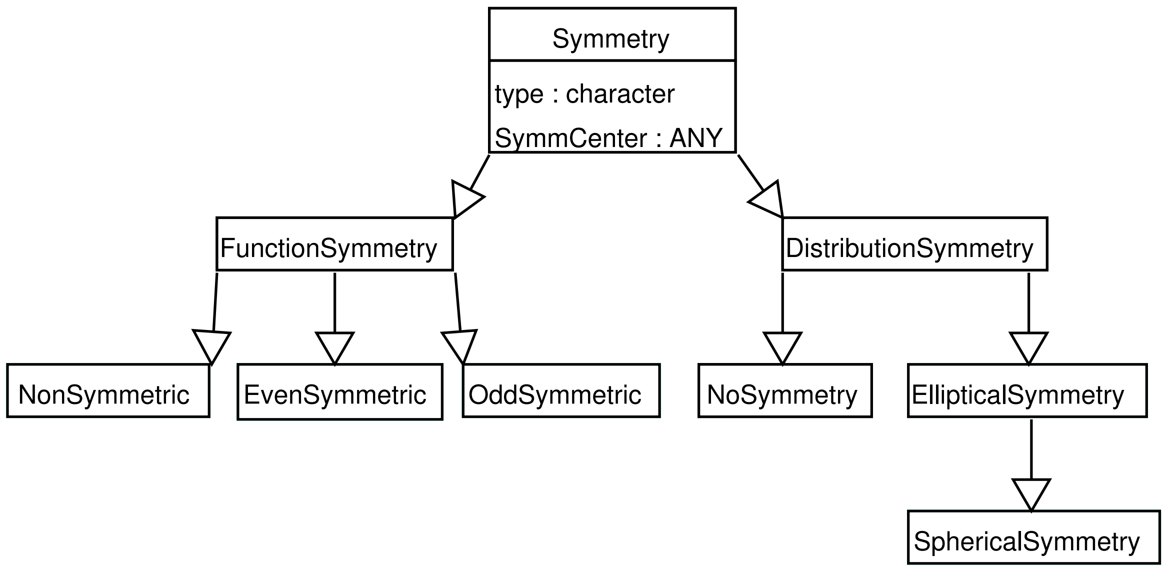
```
    ### props:
[1] "The normal location and scale family is invariant under"
[2] "the group of transformations 'g(x) = sd*x + mean'"
[3] "with location parameter 'mean' and scale parameter 'sd'"
```

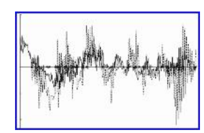
4 (d) Semi-symbolic calculus

- Situation:
  - we have a certain abstract property for our model (e.g. symmetry)
  - whether this property holds or not cannot be decided (exactly) on basis of numeric evaluations (e.g. convergence?)
  - as a logical statement we can “calculate” with this property and even deduce further properties
  - important for evaluation of high dimensional integrals



- Approach
  - in classical (linear) hierarchical inheritance relations of objects: not clear in which order we should inherit abstract properties. . .
  - introduce symbolic/logical flags as members(slots) of objects and interfere into dispatching mechanism. . .
- Realization

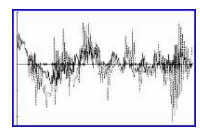




Peter Ruckdeschel
Matthias Kohl

5 Examples of optimally robust estimation

1. Estimation of location and scale
   - `X` a contaminated sample from $\mathcal{N}\,(\mathtt{mean}, \mathtt{sd}^2)$
   - want to estimate location and scale
   - next steps see `R`-Example
2. Generation of a new L_2-differentiable family:
   - censored Poisson distribution with parameter $\lambda > 0$, i.e. we only observe realizations > 0
   - next steps see `R`-Example



6 Summary/Outlook

6 (a) Summary

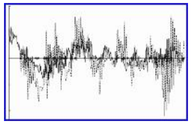
covered so far:

- computation of optimal ICs for all(!) L_2-diff'ble models based on univariate distributions
- Kolmogorov minimum distance estimator as starting estimator
- provide optimally robust estimators by means of one-step constructions

6 (b) Open issues – some perhaps settled at rsr?

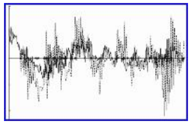
1. use of S-classes for model formula $\rightsquigarrow$ `rlm` extending `lm` also available for infinitesimal robustness
2. better and standardized user-interfaces
3. (more) standardized output
4. use of other robust diagnostic plots. . .
5. reporting: use of `XML` for the storage of meta-information about generated objects
6. use of package `Matrix`
7. one generic method for `ksEstimator`
8. further functionals for `distrEx`

Peter Ruckdeschel
Matthias Kohl



9. extension of class `RiskType`: getRiskIC
10. mStepEstimator `m = Inf`: $\hat{=}$ iteration until “convergence”
11. further distances
12. better use of symmetry and group invariances
13. special group generic for invertible operators for the exact determination of image distributions
14. `liesInSupport`: allow for logical operations for slot `'img'` of distributions
15. Lower case for Dimension > 1

. . . many more



References

[1] H. Bengtsson. The R.oo package - object-oriented programming with references using standard R code. In K. Hornik, F. Leisch, and A. Zeileis, editors, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, Vienna, Austria, March 2003. Published as `http://www.ci.tuwien.ac.at/Conferences/DSC-2003/` .

[2] J. M. Chambers. *Programming with data. A guide to the S language*. Springer, 1998. `http://cm.bell-labs.com/stat/Sbook/index.html`.

[3] —— and T. J. Hastie. *Statistical Models in S.*. Chapman & Hall, 1992.

[4] M. K. *Numerical contributions to the asymptotic theory of robustness*. Dissertation, Universität Bayreuth, Bayreuth, 2005. Available under `http://stamats.de/ThesisMKohl.pdf`

[5] ——, P. R., and T. Stabla. General Purpose Convolution Algorithm for Distributions in S4-Classes by means of FFT. Technical report, Feb. 2005. Also available under `http://www.uni-bayreuth.de/departments/math/org/ /mathe7/RUCKDESCHEL/pubs/comp.pdf`.

[6] P. R., M. K., T. Stabla, and F. Camphausen. `S4` *Classes for Distributions— a manual for package* `distr` version 1.5, March 2005. `http://www.uni-bayreuth.de/departments/math/org/mathe7/DISTR/distr.pdf`.

[7] ——. `S4` Classes for Distributions. Submitted. March 2005. Also available in `http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/distr-rnews.pdf`.

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