R-packages for infinitesimal robustness

M. Kohl and P. Ruckdeschel[1]

[1] Mathematik VII, University of Bayreuth, D-95440 Bayreuth, Germany

Keywords: infinitesimal robustness, R, S4-classes for distributions, S4-classes for random variables, optimal-robust influence curves

We present some R-packages designed for a conceptual adaptation of infinitesimal robustness theory with both slides and "online" in R.

1 R Package distr

Package distr provides a conceptual treatment of univariate distributions by means of S4–classes (cf. Chambers (1998)). A mother class is introduced with slots for a parameter and – most important – for the four constitutive methods r, d, p, and q for simulation, respectively for evaluation of density / c.d.f. and quantile function of the corresponding distribution. All distributions of R's base package are implemented as subclasses.
By means of these classes, we may automatically generate new distribution objects with corresponding r, d, p, and q-slots for the laws of r.v.'s under standard mathematical univariate transformations and under convolution of independent r.v.'s. For distribution objects X and Y expressions like 3*X+sin(exp(-Y/4+3)) have their natural interpretation as corresponding image distributions (where X and Y are stochastically independent).

2 R Package distrEx

Package distr is extended by package distrEx which provides some additional features like evaluation of certain functionals on distributions like expectation, variance, median, and also some distances between distributions like total variation-, Kolmogorov and Hellinger-distance (and "convex-contamination-distance"). As an example, for distribution objects X and Y, E(X) gives the expectation of X and KolmogorovDistance(X, Y) the Kolmogorov distance between X and Y.

3 R Package RandVar

In this package, we implement random variables by means of S4–classes: A random variable consists of a list of functions and slots for domain and range of this variable. We provide various mathematical operations for random variables. For instance, 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.

4 R Package ROptEst

The R packages distr, distrEx and RandVar are used to implement S4 classes and methods for optimally robust estimation in the sense of Rieder (1994). That is, we consider L_2 differentiable parametric models in the framework of infinitesimal (shrinking at a rate of $\sqrt{n}$) neighborhoods. By means of the packages distr, distrEx and RandVar we are able to implement **one** algorithm which works for a whole class of various models, thus avoiding redundancy and simplifying maintenance

of the algorithm.
So far, we cover computation of optimal influence curves for all(!) L_2 differentiable parametric families which are based on a univariate distribution. With the Kolmogorov minimum distance estimator serving as starting estimator we provide optimally robust estimators by means of one-step constructions.
For example, let `X` be some contaminated sample where we want to estimate location and scale and assume the ideal model distribution is $\mathcal{N}(\texttt{mean}, \texttt{sd}^2)$. Then, we can proceed as follows. We first determine the initial estimate via

```
> est <- ksEstimator(x = X, distribution = Norm())
```

Then, we define the ideal model

```
> N <- NormLocationScaleFamily(mean = est[1], sd = est[2])
```

and generate the corresponding infinitesimal robust model for contamination neighborhoods of radius `r`

```
> RobN <- InfRobModel(center = N,
+                     neighbor = ContNeighborhood(radius = r))
```

Finally, we compute the corresponding MSE optimal influence curve

```
> IC <- optIC(model = RobN, risk = asMSE())
```

and determine the optimally robust estimator by means of one-step construction

```
> estRob <- oneStepEstimator(x = X, IC = IC, start = est)
```

References

H. Bengtsson (2003). The R.oo package - Object-Oriented Programming with References Using Standard R Code. In K. Hornik, F. Leisch and A. Zeileis (eds.): *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. March 20–22, Vienna, Austria. Available under: `http://www.ci.tuwien.ac.at/Conferences/DSC-2003/`

J. Chambers (1998). Programming with data: a guide to the `S` language. Springer.

M. Kohl (2005). *Numerical contributions to the asymptotic theory of robustness*. Dissertation, Universität Bayreuth, Bayreuth.

R Development Core Team (2005). R: A language and environment for statistical computing. `http://www.r-project.org`.

H. Rieder (1994). *Robust asymptotic statistics*. Springer.

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