"RobExtremes": Robust Extreme Value Statistics — a New Member in the RobASt-Family of R Packages

Peter Ruckdeschel1,2,⋆, Matthias Kohl3, Nataliya Horbenko4

1. Fraunhofer ITWM, Kaiserslautern, Germany
2. Kaiserslautern University, Germany
3. Furtwangen University, Germany
4. KPMG
⋆Contact author: peter.ruckdeschel@itwm.fraunhofer.de

Keywords: robustness, extreme value statistics, diagnostics, operational risk, hospital length of stay

As a software offspring of VW-foundation funded project "Robust Risk Estimation", we present R package "RobExtremes", covering the scope of our project and building up on the framework of the distrXXX and RobAStXXX families of packages implemented and maintained—together with varying coauthors—by the present authors.

Starting with package "distr", providing an object-oriented framework for probability distributions, and extended by packages "distrEx" and "RandVar" with functionals and random variables acting on these distribution objects, we have set up an arithmetics for probability models. In particular, in package "distrMod", we have built up an infrastructure for general smooth parametric models such that you can write, e.g. MLEstimator(data, model).

In a further step, this approach has then been extended to cover the setup of infinitesimally robust statistics as presented in detail in Kohl et al. (2010). The corner stones are packages "RobAStBase" and "ROptEst", the former including general concepts of robust statistics such as influence curves and corresponding diagnostics, the latter general infrastructure for optimally-robust estimators, respectively.

In the RobAStXXX family of packages, we have implemented optimally robust estimation in the infinitesimal setup of Rieder (1994), i.e., $L_2$-differentiable parametric models, shrinking neighborhoods, etc. By our general approach we may employ one algorithm for a large class of probability models, thus avoiding redundancy and simplifying maintenance.

Package "RobExtremes" implements the general LD estimators introduced in Marazzi and Ruffieux (1999), in particular including the high-breakdown point estimators medSn, medQn, and medkMAD discussed in Ruckdeschel and Horbenko (2012). In addition, an interpolation technique is applied to speed-up computation of the optimally-robust estimators MBRE, OMSE, RMXE.

We demonstrate this, together with corresponding diagnostics at some real data sets from the context of hospital length of stay and operational risk of a bank, as considered within our project.

References