distrMod — an S4-class based package for statistical models

P. Ruckdeschel[1] and M. Kohl[2]

[1] Fraunhofer Institut für Techno– und Wirtschaftsmathematik ITWM, Abteilung Finanzmathematik, Fraunhofer-Platz 1, D-67663 Kaiserslautern
[2] Universität Bayreuth, Mathematisches Intstitut, D-95440 Bayreuth

The S4 concept ([1]) is a strong tool for writing unified algorithms. As an example for this in R ([3]), we present a new package distrMod for a conceptual implementation of statistical models based on these S4-classes. It is part of the distrXXX-family of packages ([4]), which is available on CRAN for quite a while, and which is developed under the infrastructure of R-Forge ([6]) in project distr ([5]).

The infrastructure to package distrMod is laid in packages distr and distrEx.

In package distr, we introduce S4 classes for distributions with slots for a parameter and for functions r, d, p, and q corresponding to functions like rnorm, dnorm, pnorm and qnorm.
We have made available quite general arithmetical operations to our distribution objects, generating new image distributions automatically, including affine transformations, standard mathematical univariate transformations like sin, abs, and convolution.

Package distrEx provides additional features like evaluation of certain functionals on distributions like expectation, variance, median, and also distances between distributions like total variation-, Hellinger-, Kolmogorov-, and Cramér-von-Mises-distance. Also, (factorized) conditional distributions and expectations are implemented.

Package distrMod then implements parametric resp. L_2 differentiable models, introducing S4-classes ParamFamily and L2ParamFamily. Based on these, quite general "Minimum Criterium"-estimators such as Maximum-Likelihood- and Minimum-Distance-Estimators are implemented.

This implementation goes beyond the scope of fitdistr from MASS ([7]), as we may work with distribution objects themselves and have available quite general expectation operators...
In short, we are able to implement **one** static algorithm which by tt S4 method dispatch may take care dynamically about various models, thus avoiding redundancy and simplifying maintenance.

This approach is also taken up to implement optimally robust estimation in the infinitesimal setup of ([4]) and its refinements in ([2]); this will be the topic of a contribution to this conference by the second author.

References

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

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

[3]. R Development Core Team (2008). R: A language and environment for statistical computing. http://www.r-project.org.

[4]. H. Rieder (1994). *Robust asymptotic statistics*. Springer.

[5]. P. Ruckdeschel, M. Kohl, T. Stabla, and F. Camphausen (2006). S4 Classes for Distributions. *R-News*, **6**(2): 10–13. http://CRAN.R-project.org/doc/Rnews/Rnews_2006-2.pdf.
Also available in extended and updated version as vignette package distrDoc on CRAN

[6]. P. Ruckdeschel, M. Kohl, T. Stabla, F. Camphausen, E. Feist, and K. Owzar (2008). Project distr. Project page on R-Forge, http://r-forge.r-project.org/projects/distr/

[7]. S. Theussl (2007) R-Forge User's Manual. http://r-forge.r-project.org/R-Forge_Manual.pdf

[8]. W. N. Venables and B. D. Ripley (2002). *Modern Applied Statistics with S*. Fourth edition. Springer.