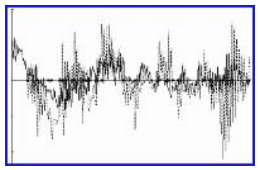
Lehrstuhl
Mathematik VII



Mathematik VII
Uni Bayreuth

S4 Classes for Distributions

UseR Conference
Vienna, May 2004

Peter Ruckdeschel
Matthias Kohl

Project realized together with
Thomas Stabla
Florian Camphausen

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

1 Package "distr"

1.1 Motivation

- I have an algorithm / program which works similarly for any distribution
- How to pass a distribution as an argument?
- Construction up to now:
  - a lot of distributions implemented to R — Gaussian, Poisson, Exponential, Gamma, etc.
  - for each:
    - cdf [$\hat{=}$p]
    - density / probability function [$\hat{=}$d]
    - quantile function [$\hat{=}$q]
    - function to simulate r.v.'s [$\hat{=}$r]
  - Naming convention: <prefix><Name>

UNIVERSITÄT BAYREUTH
Mathematik VII
Uni Bayreuth

Peter Ruckdeschel
Matthias Kohl

S4-Classes for Distributions

May 2004

- e.g. to get the median of a general distribution:

```
mymedian ← function(vtlg, ...)
{ eval(parse(text=
        paste("x←q", vtlg,
              "(1/2,...)", sep="")))
  return(x)}
```

- better: having a "variable type" distribution and functions `p`, `d`, `q`, `r` defined for this type
- then: **q**(x) returns the quantile function $\rightsquigarrow$
  median ←**function**(X , ...){ **q**(X)(0.5,...)}

$\Longrightarrow$ Development of this concept in a new package "`distr`"

UNIVERSITÄT BAYREUTH
Mathematik VII
Uni Bayreuth

Peter Ruckdeschel
Matthias Kohl

S4-Classes for Distributions

May 2004

1.2 Designing Principles

- open, easily extensible by any volunteer in the `R` community
- preserve naming and notation from `R-base`
- suggestive notation for automatically generated methods `r`, `d`, `p`, and `q`

Peter Ruckdeschel
Matthias Kohl

2 Organization in classes

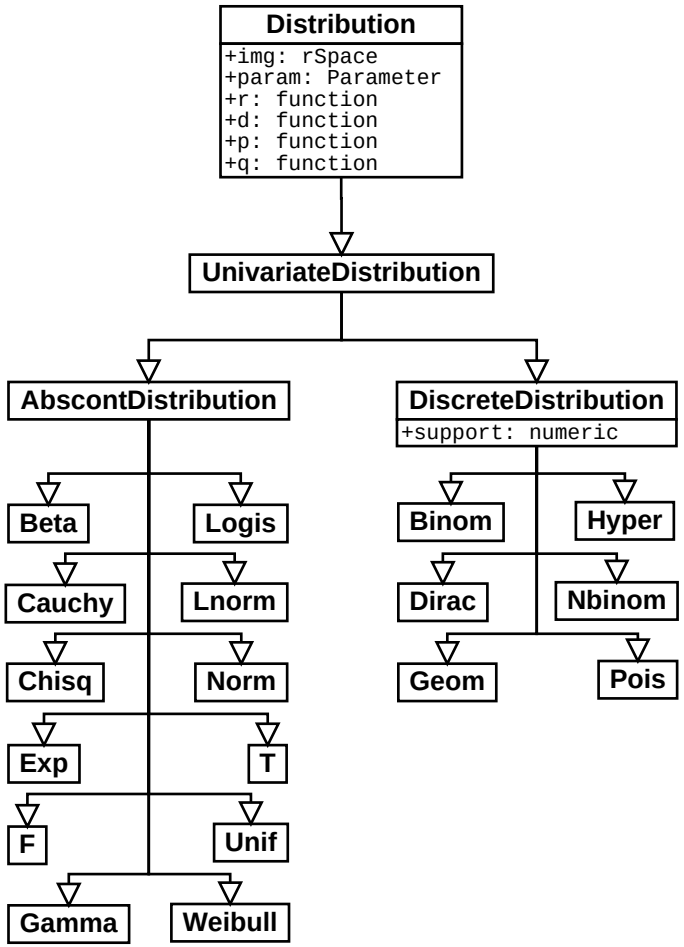
Distribution classes: Implementation of the concept of random variables / distributions in R. Mother class is the virtual class `Distribution` which (up to now) has the subclass `UnivariateDistribution`. The class of univariate distributions is divided up in absolutely continuous distributions with class `AbscontDistribution` and discrete distributions with class `DiscreteDistribution`.

Parameter classes: Classes for the parameters of distributions.

Simulation classes: Gather all information about a simulation in one object of class `DataClass`; has subclasses `Simulation` and `Contsimulation` (simulation of contaminated data).

Evaluation class: Evaluation of a new procedure (e.g. an estimator) on large set of simulation and storing of all relevant information in an object of class `Evaluation`.

Peter Ruckdeschel
Matthias Kohl



slot param is filled with an object
of the corresponding Parameter class

UNIVERSITÄT BAYREUTH

Mathematik VII
Uni Bayreuth

Peter Ruckdeschel
Matthias Kohl

S4-Classes for Distributions

May 2004

3 Possible Operations

Affine linear transformation: Expressions like Y←(3∗X+5)/4 for an object X of class `AbscontDistribution` or `DiscreteDistribution` giving an object Y of class `AbscontDistribution` or `DiscreteDistribution`.

Unary mathematical operations included in the group math are available so that expressions like Y←**exp**(**sin**((3∗ X+5)/4)) are admitted.

Convolution of probability distributions is done via the exact formulas where available, respectively, via explicit calculations. For absolutely continuous summands we implemented an accurate convolution method by means FFT (c.f. Kohl et al. (04)).

UNIVERSITÄT BAYREUTH
Mathematik VII
Uni Bayreuth

Peter Ruckdeschel
Matthias Kohl

S4-Classes for Distributions

May 2004

4 Availability

- Homepage on
  `http://www.uni-bayreuth.de/`
  `/departments/math/org/mathe7/DISTR`
- on CRAN soon ...—
  `http://cran.r-project.org/mirrors.html`
- once again: Homepage for R: `http://www.r-project.org/`

UNIVERSITÄT BAYREUTH
Mathematik VII
Uni Bayreuth

Peter Ruckdeschel
Matthias Kohl

S4-Classes for Distributions

May 2004

References

Chambers J.M. (1998): *Programming with data. A guide to the S language.*. Springer.
`http://cm.bell-labs.com/stat/Sbook/index.html`. 4

Gentleman R. (2003): *Object Orientated Programming. Slides of a Short Course held in Auckland*.
`http://www.stat.auckland.ac.nz/S-Workshop/Gentleman/Methods.pdf`. 4

Kohl M., Ruckdeschel P. and Stabla T. (2004): Computation of the convolution of arbitrary probabilities by means of FFT. unpublished manuscript. 4

Rice J.A. (1988): *Mathematical statistics and data analysis.*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software., Pacific Grove, California. 4

Ruckdeschel P., Kohl M., Stabla T. and Camphausen F. (2004): `S4` *Classes for Distributions*.
`http://www.uni-bayreuth.de/departments/math/org/mathe7/DISTR`. 4

UNIVERSITÄT BAYREUTH

Mathematik VII
Uni Bayreuth

Peter Ruckdeschel
Matthias Kohl

S4-Classes for Distributions

May 2004