Abstract

In this vignette, we give short examples how to produce new distributions in packages "distr" and "distrEx". This vignette refers to package version 2.7.

Basically there are three ways to produce new distributions in packages "distr" and "distrEx":

1. automatic generation of single distribution objects by arithmetics and the like
2. using generating functions to produce single distribution objects
3. defining new distribution classes / doing it from scratch

We will give short examples of all three of them.

*University Oldenburg, Oldenburg
†FH Furtwangen
1 Automatic generation by arithmetics and the like

We have made available quite general arithmetical operations to our distribution objects, generating new image distribution objects automatically. As an example, try

```r
require(distr)
N <- Norm(mean = 2, sd = 1.3)
P <- Pois(lambda = 1.2)
Z <- 2*N + 3 + P
Z

## Distribution Object of Class: AbscontDistribution

## Warning in methods::show(x): arithmetics on distributions are understood as operations on r.v.'s
## see ’distrARITH()’; for switching off this warning see ’?distroptions’

plot(Z, panel.first = grid(), lwd=3)
```
\textbf{Density of AbscontDistribution} \hspace{2cm} \textbf{CDF of AbscontDistribution} \hspace{2cm} \textbf{Quantile function of AbscontDistribution}

\textbf{p(Z)(0.4)}

\#
\[ \text{[1]} 0.002415387 \]

\textbf{q(Z)(0.3)}

\#
\[ \text{[1]} 6.705068 \]

\textit{\# in RStudio or Jupyter IRKernel, use q.l(.)(.) instead of q(.)(.)}

\textbf{Zs <- r(Z)(50)}

\textbf{Zs}

\#
\[ \text{[1]} 9.6026912 7.5092843 9.2283725 7.4998820 8.5378200 7.1567203 \]
\[ \text{[7]} 5.8395905 8.1853107 13.1027758 12.3038700 11.2310660 15.1092333 \]
\[ \text{[13]} 13.6715143 10.6412671 11.7288400 6.9196626 6.5346967 5.6006608 \]
\[ \text{[19]} 5.9584914 9.1192134 5.6846415 10.3036550 6.9125849 6.0939167 \]
\[ \text{[25]} 0.5189119 8.6846535 8.5178985 6.7225699 9.5996720 11.5000103 \]
Comment:
Let \( N \) an object of class "Norm" with parameters \( \text{mean}=2, \text{sd}=1.3 \) and let \( P \) an object of class "Pois" with parameter \( \lambda=1.2 \). Assigning to \( Z \) the expression \( 2 \times N + 3 + P \), a new distribution object is generated —of class "AbscontDistribution" in our case— so that identifying \( N, P, Z \) with random variables distributed according to \( N, P, Z \), \( \mathcal{L}(Z) = \mathcal{L}(2 \times N + 3 + P) \), and writing \( p(Z)(0.4) \) we get \( P(Z \leq 0.4), q(Z)(0.3) \) the 30%-quantile of \( Z \), and with \( r(Z)(50) \) we generate 50 pseudo random numbers distributed according to \( Z \), while the \texttt{plot} command generates the above figure.

In the environments of RStudio, see \url{https://www.rstudio.com/} and Jupyter IRKernel, see \url{https://github.com/IRkernel/IRkernel}, calls to \texttt{q} are caught away from standard R evaluation and are treated in a non-standard way. This non-standard evaluation in particular throws errors at calls to our accessor methods \texttt{q} to slot \texttt{q} of the respective distribution object. To amend this, from version 2.6 on, we provide function \texttt{q.l} (for left-continuous quantile function) as alias to our accessors \texttt{q}, so that all our package functionality also becomes available in RStudio and IRKernel.

There are caveats to take care about; for details refer to the (larger) vignette \texttt{distr} in package "distrDoc".

2 Using generating functions

If you want to generate a single distribution object (without any particular parameter) generating functions are the method of choice:

Objects of classes \texttt{LatticeDistribution} resp. \texttt{DiscreteDistribution}, \texttt{AbscontDistribution}, may be generated using the generating functions \texttt{LatticeDistribution()} resp. \texttt{DiscreteDistribution()} resp. \texttt{AbscontDistribution()}; see also the corresponding help.

E.g., to produce a discrete distribution with support \((1, 5, 7, 21)\) with corresponding probabilities \((0.1, 0.1, 0.6, 0.2)\) we may write

\[
D \leftarrow \text{DiscreteDistribution}\left(\text{supp} = c(1,5,7,21), \text{prob} = c(0.1,0.1,0.6,0.2)\right)
\]

\[
D
\]

\#
\#
\#
\#
## Distribution Object of Class: DiscreteDistribution

\[
\text{plot}(D, \text{panel.first} = \text{grid(lwd=2)}, \text{lwd} = 3)
\]
and to generate an absolutely continuous distribution with density proportional to $e^{-|x|^3}$, we write

```r
AC <- AbscontDistribution(d = function(x) exp(-abs(x)^3), withStand = TRUE)
AC
```

```
## Distribution Object of Class: AbscontDistribution
```

```r
plot(AC, panel.first = grid(lwd=2), lwd = 3)
```
3 Doing it from scratch

If you would like to create new parametric distributions, using already implemented \( r, d, p, \) and \( q \) functions (e.g. implementing additional distributions realized in another CRAN package), you should probably envisage introducing new distribution S4 (sub-)classes and hence better look at the implementation of some discrete and continuous parametric distribution classes in package "distr". Hint: download the .tar.gz file; extract it to some temp folder; look at subdirectories R and man.
The general procedure is as follows

1. introduce a new subclass of class `Parameter`

2. introduce a new subclass of `LatticeDistribution`/`DiscreteDistribution` (if discrete) or of class `AbscontDistribution` (if continuous).

3. define accessor and replacement functions for the “slots” of the parameter (e.g. ”size” and ”prob” in the binomial case), possibly with new generics

4. (possibly) define a validity function

5. define a generating function

6. if existing, define particular convolution methods or similar particular methods for this new distribution class

7. create .Rd files for the
   - parameter class
   - distribution class

8. if analytic expressions are available, define particular `E`, `var`, `skewness`, and `kurtosis`-methods and if so, also document\(^1\) the corresponding methods in the distribution class .Rd file

Let’s go through the steps in the example case of the Binomial implementation in packages "distr" and "distrEx":

1. in "distr", see source in R/AllClasses.R,

   ```r
   ## Class: BinomParameter
   setClass("BinomParameter",
             representation = representation(size = "numeric", prob = "numeric"),
             prototype = prototype(size = 1, prob = 0.5, name =
                           gettext("Parameter of a Binomial distribution"),
             contains = "Parameter"
           )
   ```

2. in "distr", see source in R/AllClasses.R,

   ```r
   ## Class: binomial distribution
   setClass("Binom",
             prototype = prototype(
               r = function(n) rbinom(n, size = 1,prob = 0.5)
             ),
           )
   ```

\(^1\)this is new, because so far, all `E`, `var`, `skewness`, and `kurtosis`-methods for “basic” distributions are documented in the "distrEx" documentation to `E`, `var`, . . . , but this would not be operational any longer for new derived classes, possibly defined in other, new packages
d = function(x, log = FALSE) {
    dbinom(x, size = 1, prob = 0.5, log = log)
},
p = function(q, lower.tail = TRUE, log.p = FALSE) {
    pbinom(q, size = 1, prob = 0.5,
    lower.tail = lower.tail, log.p = log.p)
},
q = function(p, lower.tail = TRUE, log.p = FALSE) {
    qbinom(p, size = 1, prob = 0.5,
    lower.tail = lower.tail, log.p = log.p)
},
img = new("Naturals"),
param = new("BinomParameter"),
support = 0:1,
lattice = new("Lattice",
    pivot = 0, width = 1, Length = 2, name = gettext(
        "lattice of a Binomial distribution"
    )
),
.logExact = TRUE,
.lowerExact = TRUE
),
contains = "LatticeDistribution"
)

3. in "distr", see source in R/BinomialDistribution.R,

```
## Access Methods
setMethod("size", "BinomParameter", function(object) object@size)
setMethod("prob", "BinomParameter", function(object) object@prob)

## Replace Methods
setReplaceMethod("size", "BinomParameter",
    function(object, value) { object@size <- value; object})
setReplaceMethod("prob", "BinomParameter",
    function(object, value) { object@prob <- value; object})
```

and R/AllGenerics,

```
if(!isGeneric("size"))
    setGeneric("size", function(object) standardGeneric("size"))
if(!isGeneric("prob"))
    setGeneric("prob", function(object) standardGeneric("prob"))
```

4. in "distr", see source in R/BinomialDistribution.R,
```r
setValidity("BinomParameter", function(object){
    if(length(prob(object)) != 1)
        stop("prob has to be a numeric of length 1")
    if(prob(object) < 0)
        stop("prob has to be in [0,1]")
    if(prob(object) > 1)
        stop("prob has to be in [0,1]")
    if(length(size(object)) != 1)
        stop("size has to be a numeric of length 1")
    if(size(object) < 1)
        stop("size has to be a natural greater than 0")
    if(!identical(floor(size(object)), size(object)))
        stop("size has to be a natural greater than 0")
    else return(TRUE)
})
```

Class "BinomParameter" [in ".GlobalEnv"]

Slots:
Name: size prob name
Class: numeric numeric character
Extends: Class "Parameter", directly Class "OptionalParameter", by class "Parameter", distance 2

5. in "distr", see source in R/BinomialDistribution.R,

```r
Binom <- function(size = 1, prob = 0.5) new("Binom", size = size, prob = prob)
```

6. in "distr", see source in R/BinomialDistribution.R,

```r
## Convolution for two binomial distributions Bin(n1,p1) and Bin(n2,p2)
## Distinguish cases
## p1 == p2 und p1 != p2

setMethod("+", c("Binom","Binom"),
    function(e1,e2){
        newsise <- size(e1) + size(e2)

        if(isTRUE(all.equal(prob(e1),prob(e2))))
            return(new("Binom", prob = prob(e1), size = newsise,
                      .withArith = TRUE))

        return(as(e1, "LatticeDistribution") + e2)
    })
```

7. in "distr", see sources in
The parameter of a binomial distribution, used by \code{Binom-class}

Objects can be created by calls of the form
\code{new("BinomParameter", prob, size)}.

Usually an object of this class is not needed on its own, it is generated automatically when an object of the class \code{Binom} is instantiated.

Objects can be created by calls of the form
\code{new("BinomParameter", prob, size)}.

```
W <- new("BinomParameter", prob=0.5, size=1)
```
This object is by Objects can be created <size(W) # size of this distribution is 1.
size(W) <- 2 # size of this distribution is now 2.

}\{distribution\}
\concept{parameter}
\concept{Binomial distribution}
\concept{S4 parameter class}

\item \man/Binom-class.Rd
\name{Binom-class}
\docType{class}
\alias{Binom-class}
\alias{Binom}
\alias{initialize.Binom-method}
\title{Class "Binom"}
\description{The binomial distribution with \code{size \eqn{=} n}, by default \eqn{=} 1, and \code{prob \eqn{=} p}, by default \eqn{=} 0.5, has density \eqn{p(x) = \binom{n}{x} \propto p^x (1-p)^{n-x}}\}
    \begin{itemize}
      \item \eqn{p(x) = \binom{n}{x} \propto p^x (1-p)^{n-x}}
      \item \eqn{p(x) = \binom{n}{x} \propto p^x (1-p)^{n-x}}
    \end{itemize}
for \eqn{x = 0, \ Idots , n}.

C.f. \code{\link{stats:Binomial}{rbinom}}
\section{Slots}{
\describe{
  \item \code{img}{Object of class \code{"Naturals"} : The space of the image of this distribution has got dimension 1 and the name "Natural_Space". }
  \item \code{param}{Object of class \code{"BinomParameter"} : the parameter of this distribution \code{\propto \code{prob}, \code{\size}}, declared at its instantiation }
  \item \code{r}{Object of class \code{"function"} : generates random numbers \code{\function \code{rbinom}}}
  \item \code{d}{Object of class \code{"function"} : \density function \code{\function dbinom}}
  \item \code{p}{Object of class \code{"function"} : \cumulative function \code{\function pbinom}}
  \item \code{q}{Object of class \code{"function"} : \inverse of the \cumulative function \code{\function qbinom}}.

The \textbf{quantile} is defined as the smallest value \( x \) such that \( F(x) \geq p \), where \( F \) is the \cumulative function. }
\item \code{support}{Object of class \code{"numeric"} : \( a \) \( \text{(sorted)} \}
\item \code{.withArith}{\code{\logical} : used internally to issue \textbf{warnings} as to \textbf{interpretation of arithmetics}}
\item \code{.withSim}{\code{\logical} : used internally to issue \textbf{warnings} as to \textbf{accuracy}}
\item \code{.logExact}{\code{\logical} : used internally to flag the \textbf{case} where there are explicit formulæ for the \textbf{log} version of \density, \textbf{cdf}, and \textbf{quantile function}}
\item \code{.lowerExact}{\code{\logical} : used internally to flag the \textbf{case} where there are explicit formulæ for the \textbf{lower} tail version of \textbf{cdf} and \textbf{quantile function}}
## DiscreteDistribution

Discrete distributions with probabilities the exact convolution formula is implemented thereby improving the general numerical accuracy.

### `initialize`

```r
initialize(e1 = "Binom", e2 = "Binom")
```

### `prob`

```r
prob()
```

### `size`

```r
size()
```

### Examples

```r
B <- Binom(prob=0.5, size=1) # B is a binomial distribution with prob=0.5 and size=1.
r(B)(1) # one random number generated from this distribution, e.g. 1
d(B)(1) # Density of this distribution is 0.5 for x=1.
p(B)(0.4) # Probability that x<0.4 is 0.5.
q(B)(.1) # q=0 is the smallest value x such that p(B)(x)>0.1.
## in RStudio or Jupyter IRKernel, use q(.)(.) instead of q(.)(.)
size(B) # size of this distribution is 1.
size(B) <- 2 # size of this distribution is now 2.
C <- Binom(prob = 0.5, size = 1) # C is a binomial distribution with prob=0.5 and size=1.
D <- Binom(prob = 0.6, size = 1) # D is a binomial distribution with prob=0.6 and size=1.
E <- B + C # E is a binomial distribution with prob=0.5 and size=3.
F <- B + D # F is an object of class LatticeDistribution.
G <- B + as(D, "DiscreteDistribution") # DiscreteDistribution
```
- you could have: `man/Binom.Rd` for the generating function; in the Binomial case, documentation is in `Binom-class.Rd`; but in case of the Gumbel distribution, in package "RobExtremes", there is such an extra `.Rd` file.

8. in "distrEx", see sources in

```r
## Loading required package: distrEx
## Extensions of Package 'distr' (version 2.8.0)
## Note: Packages "e1071", "moments", "fBasics" should be attached /before/ package "distrEx". See distrExMASK(). Note: Extreme value distribution functionality has been moved to
## package "RobExtremes". See distrExMOVED().
## For more information see ?"distrEx", NEWS("distrEx"), as well as
## http://distr.r-forge.r-project.org/
## Package "distrDoc" provides a vignette to this package as well as to several related packages; try vignette("distr").

## Attaching package: ‘distrEx’

## The following objects are masked from ‘package:stats’:
##     IQR, mad, median, var

• Expectation.R,

```r
setMethod("E", signature(object = "Binom",
    fun = "missing",
    cond = "missing"),
    function(object, low = NULL, upp = NULL, ...){
if(!is.null(low)) if(low <= min(support(object))) low <- NULL
if(!is.null(upp)) if(upp >= max(support(object))) upp <- NULL
if(is.null(low) & is.null(upp))
    return(size(object) * prob(object))
else{
    if(is.null(low)) low <- -Inf
    if(is.null(upp)) upp <- Inf
    if(low == -Inf){
        if(upp == Inf) return(size(object) * prob(object))
    else return(m1df(object, upper = upp, ...))
```
} else {
  E1 <- mldf(object, upper = low, ...)
  E2 <- if (upp == Inf)
    size(object) * prob(object) else mldf(object, upper = upp, ...)
  return(E2 - E1)
}

• Functionals.R,

  setMethod("var", signature(x = "Binom"),
  function(x, ...){
    dots <- match.call(call = sys.call(sys.parent(1)),
      expand.dots = FALSE)$"...
    fun <- NULL; cond <- NULL; low <- NULL; upp <- NULL
    if (hasArg(low)) low <- dots$low
    if (hasArg(upp)) upp <- dots$upp
    if (hasArg(fun) || hasArg(cond) || is.null(low) || is.null(upp))
      return(var(as(x, "DiscreteDistribution"), ...))
    else
      return(size(x) * prob(x) * (1 - prob(x)))
  })

• skewness.R,

  setMethod("skewness", signature(x = "Binom"),
  function(x, ...){
    dots <- match.call(call = sys.call(sys.parent(1)),
      expand.dots = FALSE)$"...
    fun <- NULL; cond <- NULL; low <- NULL; upp <- NULL
    if (hasArg(low)) low <- dots$low
    if (hasArg(upp)) upp <- dots$upp
    if (hasArg(fun) || hasArg(cond) || is.null(low) || is.null(upp))
      return(skewness(as(x, "DiscreteDistribution"), ...))
    else
      return((1 - 2 * prob(x)) / sqrt(size(x) * prob(x) * (1 - prob(x))))
  })

• kurtosis.R,

  setMethod("kurtosis", signature(x = "Binom"),
  function(x, ...){
    dots <- match.call(call = sys.call(sys.parent(1)),
      expand.dots = FALSE)$"...
    fun <- NULL; cond <- NULL; low <- NULL; upp <- NULL
    if (hasArg(low)) low <- dots$low

if(hasArg(upp)) upp <- dots$upp
if(hasArg(fun)||hasArg(cond)||is.null(low)||is.null(upp))
  return(kurtosis(as(x,"DiscreteDistribution"),...))
else
  p <- prob(x)
  return((1-6*p*(1-p))/(size(x)*p*(1-p)))
}

The procedure will be similar for any new class of distributions.

Comment  In the classes in package "distr" (historically the “oldest” in the development of this project), we still use initialize methods; this is no longer needed, if you provide generating functions; for this “more recent” approach, confer the realization of class Gumbel in package "RobExtremes".

4 Help needed / collaboration welcome

You are — as announced on http://distr.r-forge.r-project.org — very welcome to collaborate in this project! See in particular https://distr.r-forge.r-project.org/HOWTO-collaborate.txt
With this you should be able to start working.

References
