RandVar: Implementation of random variables by means of S4 classes and methods

Matthias Kohl
E-Mail: matthias.kohl@stamats.de

Abstract

In this package we implement random variables by means of S4 classes and methods.
This vignette corresponds to Appendix D.2 in Kohl (2005) [3].

Contents

1 S4 Classes ........................................ 2
2 Functions and Methods ..................... 3
3 Odds and Ends .................................. 8
1 S4 Classes

The S4 class `RandVariable` (cf. Figure 1) has the slots `Map`, `Domain` and `Range` where `Map` contains a list of functions which are measurable maps from `Domain` to `Range`. The elements contained in the list `Map` must be functions in one argument named `x`. We do not allow further parameters for these functions as this would lead to inconsistent objects. Strictly speaking, an object of class `RandVariable` would represent not only one random variable but a whole set of random variables depending on these parameters.

The slots `Domain` and `Range` are filled with an object of class `OptionalrSpace`; i.e., they contain `NULL` or an object of class `rSpace` (see package `distr` [4]). In case of `EuclRandVariable` and `RealRandVariable` the slot `Range` is filled with an object of class `Euclideanspace` and `Reals`, respectively. The class `EuclRandMatrix` additionally has the slot `Dim` which is a vector of integers and contains the dimensions of the Euclidean random matrix.

Using these S4 classes there are two possibilities to implement a $\mathbb{R}^k$ valued random variable. First, we could define a `EuclRandVariable` whose slot `Map` contains a list with one function which maps to $\mathbb{R}^k$; i.e., the slot `Range` is a $k$-dimensional Euclidean space. Second, we could define a `EuclRandVariable` whose slot `Map` contains a list with $n$ functions (projections) which map to $\mathbb{R}^m$ where $k = m \times n$. Now, the slot `Range` is an $m$-dimensional Euclidean space. Since it is sometimes convenient to regard a $\mathbb{R}^k$ valued random variable as measurable map consisting of $\mathbb{R}^{k_i}$ valued maps where $\sum k_i = k$, we introduce a further class called `EuclRandVarList`. With this class we can now define $\mathbb{R}^k$ valued random variables as a list of $\mathbb{R}^{k_i}$ valued random variables with compatible domains. More precisely, the elements of a `EuclRandVarList` may even have very different ranges (not necessarily Euclidean spaces) they only need to have compatible domains which is checked via the generic function `compatibleDomains`. 
2 Functions and Methods

As in case of package distrEx [4], we follow the advices of Section 7.3 of [1] and [2]. That is, we introduce generating functions as well as accessor and replacement functions. A short description of the implemented generating functions is given in Table 1.

While there are accessor functions for all slots of the newly defined S4 classes, replacement functions are only implemented for those slots a user should modify.

Our next goal was that one can use these classes of random variables like ordinary
### Generating Functions

<table>
<thead>
<tr>
<th>Generating Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EuclRandMatrix</td>
<td>Generates an object of class EuclRandMatrix</td>
</tr>
<tr>
<td>EuclRandVariable</td>
<td>Generates an object of class EuclRandVariable</td>
</tr>
<tr>
<td>EuclRandVarList</td>
<td>Generates an object of class EuclRandVarList</td>
</tr>
<tr>
<td>RandVariable</td>
<td>Generates an object of class RandVariable</td>
</tr>
<tr>
<td>RealRandVariable</td>
<td>Generates an object of class RealRandVariable</td>
</tr>
</tbody>
</table>

Table 1: Generating functions of package RandVar.

numeric vectors or matrices. Hence, we overloaded the S4 group generic functions Arith and Math as well as matrix multiplication %*%. For the matrix multiplication of EuclRandVarLists we additionally introduced the operator %m%. Now, if we have random variables X and Y, a numerical vector v and a numerical matrix M (with compatible dimensions) we can for instance generate

```r
> library(RandVar)
> distroptions("withSweave" = TRUE) ## only for use in Sweave - document; set to FALSE else!
> (X <- RealRandVariable(Map = list(function(x){x}, function(x){x^2}), Domain = Reals(), RanAn object of class Map(X)
```

```r
[[1]]
function (x)
{
  x
}

[[2]]
function (x)
{
  x^2
}
```

```r
> evalRandVar(X, 2)
```

```r
[,1]
[1,]  2
[2,]  4
```
> evalRandVar(X, as.matrix(seq(2, 10, 2)))

, , 1

[1,]  2  4  6  8 10
[2,]  4 16 36 64 100

> R1 <- exp(X-1)
> Map(R1)

[[1]]
function (x)
{
  f1 <- function (x)
  {
    f1 <- function (x)
    {
      x
      }
    f1(x) - 1
    }
  exp(f1(x))
}
<environment: 0x7813988>

[[2]]
function (x)
{
  f1 <- function (x)
  {
    f1 <- function (x)
    {
      x^2
      }
    f1(x) - 1
    }
  exp(f1(x))
}
<environment: 0x7813988>

> R2 <- exp(X-1:2)
> Map(R2)
[[1]]
function (x)
{
    f1 <- function (x)
    {
        f1 <- function (x)
        {
            x
        }
        f1(x) - 1L
    }
    exp(f1(x))
}
<environment: 0x787a4c0>

[[2]]
function (x)
{
    f1 <- function (x)
    {
        f1 <- function (x)
        {
            x^2
        }
        f1(x) - 2L
    }
    exp(f1(x))
}
<environment: 0x787a4c0>

> (Y <- RealRandVariable(Map = list(function(x){sin(x)},function(x){cos(x)}), Domain = Real))
An object of class

> Map(Y)

[[1]]
function (x)
{
    sin(x)
}
function (x)
{
    cos(x)
}

> R3 <- X %*% Y
> dimension(R3)
[1] 1

> #evalRandVar(R3, 2)
> 2*sin(2) + 2^2*cos(2)
[1] 0.1540075

> (R4 <- X %*% t(Y))
An object of class
dimension(R4)
[1] 4

> #evalRandVar(R4, 2)
> (M <- matrix(c(2*sin(2), 2^2*sin(2), 2*cos(2), 2^2*cos(2)), ncol = 2))

    [,1]      [,2]
[1,]  1.818595 -0.8322937
[2,]  3.637190 -1.6645873

> (R5 <- M %*% R4)
An object of class

We also implemented S4 methods for the generic function E of package distrEx [4]. That is, given some distribution D, respectively some conditional distribution CD and some random variable X we can compute the (conditional) expectation of X under D, respectively CD simply by

> D <- Norm()
> E(object = D, fun = X)

[1] 0.0000000 0.9999942
> \( E(D) \)
[1] 0

> \( \text{var}(D) \)
[1] 1

> (CD <- \text{LMCondDistribution}(\text{theta} = 1))

Distribution object of class: \text{AbscontCondDistribution}
\text{theta}: 1
\text{intercept}: 0
\text{scale}: 1
## cond:
\text{name}: conditioning by an Euclidean space
\text{Range}: Euclidean Space with dimension 1

> \( E(\text{object} = \text{CD}, \text{fun} = X, \text{cond} = 2) \)
[1] 2.000000 4.999993

> \( E(\text{Norm}(\text{mean} = 2)) \)
[1] 2

> \( E(\text{Norm}(\text{mean} = 2), \text{fun} = \text{function}(x)\{x^2\}) \)
[1] 4.999993

for some given condition \text{cond}.

In addition, we define methods for the generic function \text{show} for the classes \text{RandVariable}, \text{EuclRandMatrix} and \text{EuclRandVarList}. There are also methods for the generic functions \text{dimension} (see package \text{distr} [4]), \text{length}, \text{ncol}, \text{nrow}, \text{t} and \text{[} (cf. package \text{base}). For more details we refer to the corresponding help pages.

Finally, we introduce several new generic functions. A brief description of these functions is given in Table 2.

For more details about the full functionality of package \text{RandVar} we refer to the source code and the corresponding help pages, respectively.

### 3 Odds and Ends

The main issue is to reduce the computation time for methods using objects of class \text{RandVariable} and its subclasses as these classes play an important role in the computation of optimally robust estimators; confer Kohl (2005) [3]. In particular, we are looking for ways to increase the computation speed of \text{evalRandVar} and \text{E}. 
<table>
<thead>
<tr>
<th>Generic Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>%m%</td>
<td>matrix multiplication for EuclRandVarLists</td>
</tr>
<tr>
<td>compatibleDomains</td>
<td>test if the domains of two random variables are compatible</td>
</tr>
<tr>
<td>evalRandVar</td>
<td>evaluation of random variables</td>
</tr>
<tr>
<td>imageDistr</td>
<td>image distribution of some distribution under some random variable</td>
</tr>
<tr>
<td>numberOfMaps</td>
<td>number of functions contained in the slots Map of the members of a EuclRandVarList</td>
</tr>
</tbody>
</table>

Table 2: New generic functions of package RandVar (without accessor and replacement functions).

References


