

ContRV

—

Lucru cu Variabile Aleatoare Continue

Moroianu Theodor

Puşcasu Felix

Tudor Raluca

January 29, 2021

Contents

1	Introducere	3
2	Echipa	3
3	Dezvoltarea Pachetului	3
4	Notiuni Avansate	4
5	Exemple De Cod	4
6	Surse De Inspiratie	6
7	Dificultati Intampinate	6
8	Lista de Obiecte Definite	7
9	Referinte	8
10	Concluzii	8

1 Introducere

În aceasta documentatie vom prezenta ideea din spatele pachetului nostru, modul in care acesta a fost implementat, si cateva exemple de utilizare a pachetului.

Prin acest proiect se doreste implementarea in R a unui pachet care sa permita lucrul cu variabile aleatoare continue. Acesta include, dar nu este restrans la:

- Crearea unei variabile aleatoare continue pornind de la o distribuție cunoscută.
- Crearea unei variabile continue pornind de la PDF.
- Crearea unei variabile continue pornind de la CDF.
- Calculul diferitor masuri ale unei variabile continue.

2 Echipa

Echipa noastra este compusa din trei membri:

1. **Moroianu Theodor** – liderul echipei
2. **Puşcasu Felix**
3. **Tudor Raluca**

3 Dezvoltarea Pachetului

Echipa s-a format cu cateva zile dupa anuntarea existentei proiectului, moment in care ne-am decis ce workflow sa folosim (*GitHub* – cu dezvoltare bazata pe *Pull-Request*-uri).

Dezvoltarea s-a facut predominant in *Linux*, si am decis sa folosim *VSCode* ca editor, multumita integrarii cu *Git*.

Pentru a construi pachetul propriu zis, am implementat un script care sa apeleze comenzile necesare de *R*.

4 Notiuni Avansate

Pentru implementarea pachetului, am folosit mai multe notiuni avansate ale limbajului *R*:

- Crearea functiilor *on-the-fly*, si imbinarea acestora in logica pachetului.
- Lucrul cu functia *integrate*, si construirea unui nivel de abstractizare pentru vectorizarea functiilor.
- Supraîncarcarea funcțiilor *print*, respectiv *plot* așa încat să respecte logica proiectului.
- Supraîncarcarea operatorilor $+$, $-$, respectiv $*$.

Am întâmpinat, de asemenea, noțiuni avansate matematice, precum:

- Adunarea, scaderea si inmutirea cu scalari ale variabilelor continue.
- Adunarea, scaderea a doua variabile aleatoare continue, pe care am rezolvat-o cu diferite convolutii.
- Extragerea unor valori dintr-o distributie, pe care am rezolvat-o prin gasirea unei bijectii pastrand masura intre intervalul $[0, 1]$ si distributie.

5 Exemple De Cod

Pentru a construi variabilele aleatoare se folosesc mai multe nivele de abstractizare, cel mai jos fiind functiile *BuildFromCDF* si *BuildFromPDF*.

Implemetarea functiei *BuildFromPDF* este urmatoarea:

```
# Creates a CoreVariable from a PDF.
#' @export
BuildFromPDF <- function(pdf) {
  vectorized_pdf <- MakeVectorized(pdf)

  cdf <- function(x) {
    ans <- c()
    for (elem in x)
      ans <- c(ans, Integrate(vectorized_pdf, -Inf, elem))
    return(ans)
  }

  variable <- CoreVariable(cdf=MakeVectorized(cdf), pdf=vectorized_pdf)

  return(variable)
}
```

Un alt exemplu interesant este extragerea unor valori aleatoare dintr-o distribuție, pe care am implementat-o cu ajutorul unei căutări binare:

```
# Genereaza N valori dintr-o repartie data de un CoreVariable
#' @export
SamplePointsFromDistribution = function(dist, size) {
  ans = c()
  val = runif(size)

  for (x in val) {
    # Cautam binar raspunsul
    p <- -100
    pas <- 100
    for (idk in 1:20) {
      act <- p + pas
      if (dist@cdf(act) <= x)
        p <- p + pas
        pas <- pas / 2
    }
    ans <- c(ans, p)
  }
  return(ans)
}
```

6 Surse De Inspiratie

Sursa principală de inspirație a fost pachetul *DiscreteRV*, pe care acest pachet (*ContRV*) încearcă să îl replice.

7 Dificultati Intampinate

Una dintre dificultățile întâmpinate a fost utilizarea funcției *integrate*, de care am avut nevoie pentru a rezolva multe dintre cerințe (e.g. trecerea de la PDF la CDF).

Problema acestei funcții, menționată chiar în documentația ei oficială, este *"If the function is approximately constant (in particular, zero) over nearly all its range it is possible that the result and error estimate may be seriously wrong."*

Altfel spus, pentru variabile aleatoare cu o distribuție egală cu 0 pe o mare parte din domeniu funcția *integrate* nu găsește răspunsul corect.

Pentru a rezolva această problemă, am luat două măsuri:

- Orice convoluție / integrală efectuată asupra variabilelor este intersectată cu intervalul $[-30, 30]$.
- Funcția *integrate* este apelată pe intervale de lungime maxim 10, cu un număr ridicat de subdivizii, pentru a minimiza șansa unui răspuns greșit.

8 Lista de Obiecte Definite

Lista obiectelor si functiilor definite in cadrul pachetului *ContrV* este data mai jos:

Nume	Parametrii	Descriere
<i>BuildFromPDF</i>	Funcție	Construieste o variabila aleatoare plecand de la un PDF.
<i>BuildFromCDF</i>	Funcție	Construieste o variabila aleatoare plecand de la un CDF.
<i>BuildFromCommonPDF</i>	Funcție, Numar	Obține variabila aleatoare conform proiectiei pe axa data de parametrul dat, plecand de la CDF (<i>Joint Cumulative Distribution Function</i>).
<i>BuildConditionalPDF</i>	Funcție, Numar, Numar	Obține variabila aleatoare conform proiectiei pe axa data de parametrul dat, plecand de la CDF, conditionata de o valoare data.
<i>BuildUniformDistribution</i>	Numar, Numar	Construieste o variabila aleatoare ce urmeaza o repartitie uniforma, plecand de la capetele stanga si dreapta.
<i>BuildNormalDistribution</i>	Numar, Numar	Construieste o variabila aleatoare ce urmeaza o repartitie normala, plecand de la medie si deviatia standard.
<i>BuildExponentialDistribution</i>	Numar	Construieste o variabila aleatoare ce urmeaza o repartitie exponentiala, plecand de la parametrul lambda (<i>rate parameter</i>).
<i>BuildChiSquareDistribution</i>	Numar	Construieste o variabila aleatoare ce urmeaza o repartitie chi-square, plecand de la numarul de grade de libertate (<i>number of degrees of freedom</i>). Se foloseste de functia <i>Gamma</i> .
<i>CheckIfFunctionIsPDF</i>	Funcție	Verifica daca o functie este PDF.
<i>ComputeNormalizationConstant</i>	Funcție	Gaseste constata de normalizare.
<i>SamplePointsFromDistribution</i>	Var. Aleatoare	Extrage puncte dintr-o distributie.
<i>ComputeMeanForFunc</i>	Var., Funcție	Gaseste media variabilei trecute prin functie.
<i>ComputeVarForFunc</i>	Var., Funcție	Gaseste varianta variabilei trecute prin functie.
<i>Conditional</i>	Var., Funcție, Funcție	Gaseste probabilitatea conditionata <i>P</i> .
<i>Average</i>	Var.	Extrage media variabilei aleatoare.
<i>Variance</i>	Var., Numar	Extrage varianta variabilei aleatoare.
<i>RawMoment</i>	Var., Numar	Obține momentul initial de ordin dat.
<i>CentralMoment</i>	Var., Numar	Obține momentul centrat de ordin dat.
<i>AverageProduct2d</i>	Funcție	Gaseste media produsului a doua var. continue.
<i>Covariance2d</i>	Funcție	Gaseste covarianta a doua var. continue.
<i>Corelation2d</i>	Funcție	Gaseste corelatia a doua var. continue.
<i>print</i>	Var. Aleatoare	Afiseaza variabila aleatoare data ca parametru.
<i>plot</i>	Var. Aleatoare	Afiseaza graficul distributiei variabilei aleatoare.
<i>Information</i>	Sir de caractere	Afiseaza o “fisa de sinteza” care conține informații de bază despre repartiția al carei nume a fost dat ca parametru.

9 Referinte

Andrew G. Glena, Lawrence M. Leemisb, John H. Drew *Computing the distribution of the product of two continuous random variables*
Computational Statistics & Data Analysis, 2002

10 Concluzii

Acest proiect a fost unul interesant, atat prin descoperirea unei noi tehnologii folosite de mii de statisticieni, cat si prin explorarea unor unelte si concepte matematice avansate.